



SANTINIKETAN  
LIBRARY.

Class No....3.71.27

Author No.....~~H23~~ H2

Shelf No.....

Accession No 44.63.











**EXAMINATIONS**  
**AND THEIR RELATION TO CULTURE AND**  
**EFFICIENCY**



# EXAMINATIONS AND THEIR RELATION TO CULTURE AND EFFICIENCY

By

P. J. HARTOG, C.I.E., M.A., B.Sc.

With a Speech by the late

EARL OF CROMER, O.M., G.C.B.

LONDON

CONSTABLE AND COMPANY LTD

*First published*  
*1918*

PRINTED IN GREAT BRITAIN

TO  
M. E. S.



# CONTENTS

	PAGE
INTRODUCTION . . . . .	ix
I. EXAMINATIONS IN THEIR BEARING ON NATIONAL EFFICIENCY . . . . .	I
SPEECH BY THE LATE EARL OF CROMER . . . . .	34
II. THE THEORY OF EXAMINATIONS . . . . .	49
NOTES :	
A. EXTRACTS FROM THE FOURTH RE- PORT OF THE ROYAL COMMISSION ON THE CIVIL SERVICE, 1914 . . . . .	80
B. ON THE REPORT OF THE TREASURY COMMITTEE ON CIVIL SERVICE, CLASS I. EXAMINATION (1917) . . . . .	89
C. THE CONNOTATION OF THE WORD "CULTURE" . . . . .	93
D. <i>KULTUR</i> AND CULTURE . . . . .	95
E. PROF. EDGEWORTH'S INVESTIGA- TIONS ON THE STATISTICS OF EX- AMINATIONS . . . . .	102
F. ON KNOWLEDGE-TESTS AND MEMORY- TESTS . . . . .	136
G. ON THE TERM "ORDER OF MERIT" . . . . .	138
INDEX . . . . .	141



"That man, I think, has had a liberal education who has been so trained in youth that his body is the ready servant of his will, and does with ease and pleasure all the work that, as a mechanism, it is capable of; whose intellect is a clear, cold, logic engine, with all its parts of equal strength, and in smooth working order; ready, like a steam engine, to be turned to any kind of work, and spin the gossamers as well as forge the anchors of the mind; whose mind is stored with a knowledge of the great and fundamental truths of Nature and of the laws of her operations; one who, no stunted ascetic, is full of life and fire, but whose passions are trained to come to heel by a vigorous will, the servant of a tender conscience; who has learned to love all beauty, whether of Nature or of art, to hate all vileness, and to respect others as himself.

"Such an one and no other, I conceive, has had a liberal education."—HUXLEY, *Science and Education*, p. 86.

## INTRODUCTION

IN the first paper in this little book, delivered as a lecture at the Royal Society of Arts in January 1911 (and published in a first edition by Hugh Rees), I put forward a plea, which received the powerful support of Lord Cromer, for a Royal Commission to inquire into the working of the examination system. The paper was discussed in a number of leading articles in the press. I cannot say what influence the paper itself, or Lord Cromer's views, or public comments, had on the Government. At any rate, in March 1912, a Royal Commission, presided over by Lord MacDonnell, was appointed with the following reference :

To inquire into and report on the methods of making appointments to and promotions in the Civil Service, including the Diplomatic and Consular Services, and the legal departments ;

To investigate the working and efficiency of the system of competitive examination

## INTRODUCTION

for such appointments, and to make recommendations for any alterations or improvements in that system which may appear to be advisable; and

To consider whether the existing scheme of organisation meets the requirements of the Public Service, and to suggest any modifications which may be needed therein.

In April 1914 (after three Reports of a formal nature accompanying the issue of volumes of evidence) the Commission published its Fourth Report (with a Minority Report) in a volume extending to over 150 folio pages.<sup>1</sup> The last paragraph of the Majority Report reads as follows: "In submitting this Report on the first part of our reference we desire," etc.—here follows a well-deserved tribute to the work of the Secretary.

Now the Report, as a matter of fact, deals not only with the first, but with the third part of the reference, and contains important suggestions for the reorganisation of the Civil Service. But with the second part of the reference—the working and efficiency of the examination system—the Commission neither profess to have dealt nor have they dealt, except in a most fragmentary fashion, and it is important to bring out this fact lest it be supposed that the work which the Commission were asked to do in regard to examina-

<sup>1</sup> Royal Commission on the Civil Service [Cd. 7338].

## INTRODUCTION

tions has, in fact, been carried out. I quote in Note A, pp. 80-88 below, the main passages of the Report dealing with the working of the examination system.

I have been unable to discover in the whole of the Fourth Report any contribution to our knowledge of the working of that system of examinations, or any suggestion as to its improvement. The Commission say that they propose "in due place" [see Note A, p. 81] to examine the defects attaching to the system, and suggest that they are susceptible of "some mitigation." But the examination of these defects is reserved for a further Report, of which, perhaps owing to the war, there has been no sign. In their formal Recommendations, the Commission say that examinations should be directed to testing, *inter alia*, "the formation of character"<sup>1</sup>; but they give no indication of how this difficult test is to be carried out. I have no desire to criticise the Royal Commission in this matter; I wish to point out that they have furnished no matter for criticism. It would be convenient to know whether the Commission have any intention of completing the investigation with which they were entrusted. A large body admirably representing many important interests, but not particularly well acquainted with the examination system, may well have recoiled before an expert and ungrateful task of this kind, and have rested

<sup>1</sup> See Report, p. 101, quoted in Note A, p. 88.

## INTRODUCTION

content with the weighty conclusions they have reached in regard to other matters. But it is a fact that ignorance of the real meaning and working of our examinations remains in 1917 what it was in 1911; and the plea for a comprehensive scientific investigation holds as good now as it did then.

It would, however, be wrong not to point out the very fruitful result to which one of the suggestions of the Majority Report of the Commission<sup>1</sup> has led in the shape of the Report of Mr. Stanley Leathes' Committee on Civil Service Examinations. If the recommendations of the Committee are adopted, they will mark a new era in our system of competitive examinations by the importance they will give to that part of examinations for which cramming is impossible, and they will go far to carry out certain suggestions which I have made in regard to this subject.<sup>2</sup>

The object of the second paper in this volume, originally published in *The Times Educational Supplement* of May 10th, 1917, though it covers somewhat the same ground as the first, is different.<sup>3</sup>

<sup>1</sup> See Majority Report, § 50, quoted on p. 87, below.

<sup>2</sup> For details in regard to the personnel and reference of the Committee and a summary of their recommendations, see Note B, p. 89 below.

<sup>3</sup> The paper, as actually written, though composed for another purpose, was submitted to Mr. Stanley Leathes' Committee at the request of one of the members; it was not printed quite *in extenso* in the *Educational Supplement*. I have made a few modifications of it for its republication.

## INTRODUCTION

It is intended to give a more general perspective of the subject as a whole, though in the last paragraph I have not been able to resist the temptation to appeal once more to examining bodies to tell us in clear and unmistakable terms the meaning of their various examination certificates. The paper may be regarded as a series of chapter headings for a larger work which I have had in mind but which I have not had the leisure to write. The subject, important in 1911, is even more important now. Mr. Fisher's remarkable speeches as President of the Board of Education, and his Education Bill of this year, seem to show that the Government are in earnest about educational reconstruction, that the long campaign of enlightenment and propaganda in which Mr. M. E. Sadler has taken the chief part is at last about to bear fruit. The working classes and their leaders, whose share in the government of the country is rapidly becoming of increasing—I was almost saying predominant—importance, are no less in earnest. But in dealing with examinations there is still timidity of thought and action that may go far to wreck our new-born hopes.

Let it not be said that examinations do not matter. Every teacher, every administrative educational body, knows that, whatever be the schemes and curricula, examinations do effectively control the class-rooms of our Secondary Schools, and that they control to a considerable extent our Universities also ; and no educational

## INTRODUCTION

reform which leaves our examination system as it is, will or can be a real reform. Professor John Burnet, the distinguished philosopher and Dean of the Faculty of Arts at St. Andrews, has written an interesting book on *Higher Education and the War*.<sup>1</sup> There are few pages in it, after the initial chapters on "Kultur and Humanism" that do not refer to examinations, directly or indirectly. Yet, speaking (on page 195) of classical teaching Professor Burnet declares that: "The written part of the examination can only deal successfully with that part of it which is destined to be forgotten as soon as it has served its purpose, and it can tell us next to nothing of what is to remain as a possession for ever." I do not disagree with Professor Burnet. On the contrary, I wish he had travelled with me further. And in this connection I want to plead for the retention of our English word "culture," which is in so great a danger of being ostracised through the recent degeneration of its German cousin *Kultur*.<sup>2</sup> I want the word especially because it enables me to summarise briefly my view of examinations in relation to education as a whole. Using the word as I have myself used it in the text, or as Professor A. N. Whitehead has recently defined it<sup>3</sup>

<sup>1</sup> Macmillan & Co., 1917.

<sup>2</sup> See Note C, p. 93, below.

<sup>3</sup> In his original and stimulating series of essays entitled, *The Organisation of Thought* (Williams & Norgate, 1917), p. 8 and *passim*.

## INTRODUCTION

—"Activity of thought, and receptiveness to beauty, and humane feeling"—*it exactly covers that vital part of education which cannot be tested by the ordinary written examination.* Examinations test, and can test, that part of a man's (or a woman's) education in which it is useful and necessary to think and act more or less on a model, and which I have called education for technical efficiency. Culture is that part of education which is meaningless if it is not sensitive and individual.

It is convenient to divide education into education for technical efficiency and education for culture, remembering, however, that efficiency in its widest sense implies culture also. The social fitness of an individual depends not on one of these factors alone but on both, and both must be considered in devising our new schemes of education. My complaint of reformers like Professor Burnet is that their criticism of the much-abused examination system leaves it as supremely unreformed as ever. And I urge two correlative reforms: first, that examinations should be made much more real and drastic tests of technical efficiency than they are at present; and, secondly, that a large portion of the educational field should be entirely freed from their influence. In what follows I have given my reasons for regarding many of our present examinations as insufficiently searching in matters recognised as of the first importance; but I may give an example here of



## INTRODUCTION

the use which should be made, in my judgment, of that portion of the school time-table which I hope some day to see freed from examination control altogether. I should like to see every boy and girl who intends to follow a scientific career obliged to attend classes in literature and history with no examination on these subjects, the teacher being forced either to interest his pupils or to suffer from their boredom, without any examination whip at his command, and I should like to see every boy or girl who intends to follow an "Arts" career obliged to attend classes in natural science of the same kind. What a revolution we should have in the teaching of these non-examination classes! What a blessing for the good teacher; what a stimulus for the poor one! I desire here to protest emphatically against the idea of Mr. R. W. Livingstone, expressed in his recent book,<sup>1</sup> that science is "unimaginative." It is quite true that elementary science can be extraordinarily dull and inhuman. So (as he sees) can elementary history, or Latin accidence; they are so in the hands of teachers who have neither historical training nor philosophic insight. The same holds good for science, and one is bound to admit that much of the teaching of "science" has been in the

<sup>1</sup> R. W. Livingstone—*A Defence of Classical Education*, 1916. [Some of the attacks on Mr. Livingstone's book, which seems to me in many ways admirable, are not justified by the text, in which he claims for science a great share in education (*cf.* pp. 51 *et seq.*)].

## INTRODUCTION

past all that Mr. Livingstone depicts it--uncritical, unhistorical, unscientific, unaware that our scientific description of nature is constantly changing with the progress of science, or that the achievements of science are marked by the individualities of great men as certainly as the achievements of literature. Even in mathematics, the most compelling of all sciences, the presentation of truths is not a unique presentation which imposes itself on all alike; there are many presentations possible, and our traditional methods, now so largely changing, are characterised by the individualities of the great mathematicians of the past. While it is true that in literature and in art there can be no compulsion of the individual to admire or accept anything (although *pro tanto*, we must be content to be regarded as barbarians if we refuse our admiration to certain of the "greatest" men); and, on the contrary, in science the compulsion in regard to some fundamental truths is forced on us by our physical environment, yet there is a great æsthetic and individual side in science, of which, from pure ignorance, many literary men fail to appreciate the existence. The terms "a beautiful theorem," "a beautiful experiment," "a beautiful theory," correspond to beauty as real to those who feel it as the beauty of a poem or a picture. The "wonders of science" commonly brought forward for our admiration are small things compared with the evolution of the

## INTRODUCTION

Newtonian theory of gravitation from the observations of Kepler, or of the evolution of Ampère's theory of electro-dynamics from a few simple experiments of Oersted. But the culture of science, like the culture of art, is not for the examination-room.

If the problems of examination touch the general problems of education at every point they nevertheless require separate investigation and treatment ; and it is for this investigation that I plead, and to which I hope that this little book may be regarded as a contribution. Will the new Secondary Schools Examination Council created by Mr. Fisher tread the old paths ? Or will they avail themselves of a really great and new opportunity ?

P. J. H.

August 1917.

Owing to absence abroad the author has been unable to see this book through the press, and he desires to express his sincere thanks to Professor T. P. Nunn and to Mr. Arthur Watson, who have most kindly revised the proofs for him, and also to Professor F. Y. Edgeworth for his help in the revision of Note E.

[The two following essays are based in part not only on previous publications by the writer, but also on a joint article on "Examinations" by Mr. Arthur Watson and himself in the eleventh edition of the *Encyclopædia Britannica*, which contains a fairly extensive bibliography of the subject. The author desires, further, to refer his readers to the article on Examinations signed by Prof. A. N. Whitehead as Chairman of the relevant committee of the "Education Reform Council," and published in the valuable collection of Reports entitled *Education Reform* (P. S. King & Son, 1917).]

# EXAMINATIONS AND THEIR RELATION TO CULTURE AND EFFICIENCY.

## I EXAMINATIONS IN THEIR BEARING ON NATIONAL EFFICIENCY

*(A lecture delivered at the Royal Society of Arts in  
January 1911.)*

MAY I begin with a personal explanation? What I have to say to-night is said on my own behalf alone, and no one else in the University which I have the honour to serve must be regarded as committed in any way to agreement either with my main thesis or with any incidental opinions I may express.

To avoid all possible misunderstanding I shall state my position at the outset. I do not desire to appear as counsel either for the plaintiff or for the defendant in the great controversy on examinations which has raged intermittently for more than half a century in this country, as in other countries, but rather as one of the public, deeply interested in the issues, and trying to see clearly through the heat and dust of the conflict.

## EXAMINATIONS AND

It is often said, and generally accepted, that the whole of English education is controlled by the examination system. It is well, however, in dealing with educational questions, to keep somewhere in the back of one's mind the fact that there is another kind of education than the education of the schools, the education that Mr. Samuel Weller's father gave him when he let him have the run of the streets and make shift for himself; well to remember that there is one branch of the national activity, not the least successful or important—I mean its commerce—which up to the present has been but little affected by examination-ideals; and that our agriculture and industries have been affected by them only to a minor extent. Our business, our agriculture, our industries, have to a large extent “shifted for themselves,” I will not say without the help of school teaching, but very largely without the help of examinations.

But the public demands that persons on whose services it relies, and for whose failures it cannot be compensated, as by a business man who fails to fulfil his contract—that these should produce some certificate of competency based on an examination, and often on a series of examinations beginning in childhood and prolonged into early manhood and beyond. Teachers, lawyers, doctors, dentists, engineers, architects, and the civil, naval, and military servants of the Crown, on whom the organisation and defence of the

## NATIONAL EFFICIENCY

Empire rest, must pass examination tests and in many cases their whole career is determined by such tests.

It is, therefore, a matter of national and imperial concern that this method of test should be wise in its design, certain and not capricious in its working, and, above all, that in its operations it should not damage the human material on which it sets a hall-mark.

Can it be said of English examinations at the present day that they fulfil these requirements? Are they wisely designed? Do they work with certainty and accuracy? Do they tend, on the whole, to improve the boys and girls, the men and women, whose education, whose lives and careers they so largely control? Are they really serving in the best possible way the national and imperial purposes they are supposed to serve? If these questions were put to a Referendum, I have little doubt of the reply.

While it may be conceded that there are many judicious examiners, and examining bodies of unquestioned wisdom, the body educational as a whole is suffering from the aches and pains of examination. Those who say, "Get rid of examinations altogether," offer us nothing in their place; they do not realise the situation. I spoke just now as if examinations affected us only from without; but have not examinations become, as it were, an artificial nervous system of our education, of which every movement is

## EXAMINATIONS AND

controlled either by their stimulus or by their power of inhibition, so that extirpation is an impossible task? If we cannot remove this tyrant that governs education from within, cannot we either reduce its powers or force it into beneficence? Quixotic as it may seem, I believe the task not to be impossible.

The task demands inquiry—and by inquiry I mean scientific investigation—patience, and perhaps some drastic reform later. My object to-night is to suggest, first of all, the outlines of an inquiry into our examination system, its design, its methods, and its results; and I shall use the word “examination” to mean a systematic test of knowledge, and of either special or general capacity or fitness, carried out under the authority of some public body.

First of all, the design of examinations. We cannot understand that design without one word as to their history and their original function. The origin of Western examinations—I say nothing of those Chinese examinations which date back three or four thousand years—is sufficiently well known to the historians; they are derived from the examinations of the mediæval universities. Those universities were corporations of teachers and students, analogous to the trade and craft guilds of the time. In the guilds there were apprentices, companions, and masters. No one was admitted to mastership until he had served his apprenticeship, nor, as a rule, until he had

## NATIONAL EFFICIENCY

shown that he could accomplish a piece of work to the satisfaction of the guild. The object of the University was to teach ; its scholars, bachelors, or pupil teachers, and masters or doctors corresponded roughly to the three classes established by the guilds, and the master or doctor (two terms at first equivalent) was one who, having served his apprenticeship and passed a definite technical test, had received the *licentia docendi*—the permission to teach. The candidate was required to teach in public at the examination precisely as he would be required to teach later in the schools. The examination fulfilled exactly the technical purpose for which it was interded. We are apt to think of the schoolmen and all their ways with a contempt inherited from the days of Bacon and Descartes ; but in discussing examinations of this kind, that distinguished historian of education, the late Professor Paulsen, of Berlin, asserts that they were well adapted to increase a student's alertness, his power of comprehending new ideas, and his ability quickly and surely to assimilate them to his own ; and that they did more to enable students to grasp subjects than the mute and solitary reviewing and cramming for our modern examinations can possibly do.<sup>1</sup> The limitations

<sup>1</sup> Germany has frequently in the past been held up to English reformers as a model country in which no examinations similar to those of Great Britain are held. The writers have in mind chiefly the question of University examinations, and are unaware of the elaborate system of State examinations which guard every professional career in Germany



## EXAMINATIONS AND

of the test were the limitations of the educational and philosophic ideals of the time. At their best, as I say, they did exactly what they were required to do. At their worst, says Canon Rashdall, they were less of a farce than certain English University pass examinations almost within the memory of persons now living.<sup>1</sup>

To trace the history of Western examinations from their origin to our own time we should need to follow up the ever-changing course of educational ideals with the development of modern societies through seven centuries, and the manifold actions and reactions during all that period between education ideals and examination ideals. For such a task we have not even the materials prepared; but, remembering that the original examination was a test of technical efficiency, excellently suited to its purpose, we can see how examinations have taken on two, and perhaps three, entirely new functions, and with far-reaching effects both on examinations and on education.

The first new function—the use of examinations to distinguish between candidates of different merit—was an early and almost inevitable develop-

except that of a University teacher, and are equally unaware of the severe system of school examinations, condemned by Paulsen, and dealt with by Prof. Burnet in his book on *Higher Education and the War*. See F. Paulsen, *The German Universities*, trans. by F. Thilly and W. Elwang, p. 25, and *Geschichte des gelehrten Unterrichts*, vol. ii., pp. 368 *et seq.*

<sup>1</sup> *The Universities of Europe in the Middle Ages*, by Hastings Rashdall, vol. i., p. 460 (Clarendon Press, 1895).

## NATIONAL EFFICIENCY

ment ; and yet we can see every day how the competitive idea leads us to forget the distinction between the efficiency of an examination and its difficulty. It is constantly said by examination-enthusiasts : " This examination is a superior examination to that, because it is harder to pass." Should we not rather ask : " Is the man who passes it a more useful member of society ? " In examinations professedly technical, like the original examinations, this question is of the utmost importance. Here the main question is, surely, not which candidate can do the most difficult things, solve the most Chinese of puzzles, remember the most unrememberable of formulæ, but whether the examination is helping to turn out the best teachers, doctors, engineers, industrial botanists, industrial chemists. In other examinations, like the highest Civil Service examinations, the element of pure difficulty, on which I shall have a word to say later, may appear in another light.

The second new function of examinations is a later development, and in some ways a characteristically, though not exclusively, English development—I mean their use to test not so much the efficiency of candidates as that of their schools. The external examination of schools and of school-children, begun by the College of Preceptors in 1853, and now carried out on an immense scale by the College and by other bodies, mainly University bodies, all over the United Kingdom, had

## EXAMINATIONS AND

primarily for its object to distinguish between the sheep and the goats of the private schools. The examination, as thus used, was organised primarily as a technical test, good or bad, not of the taught, but of the teacher. But are not the real interests of the taught, which it was originally intended to serve, sometimes, or perhaps often, lost sight of in the process? Here, again, the difficulty of an examination is apt to be confused with its efficiency.

There is a third function of examinations, not easy to distinguish altogether from the original function of testing efficiency—I mean their use to test what is called “general culture” In dealing with this question we find a certain amount of common agreement as to what is the best policy, combined with a powerlessness to carry out that policy in the face of our present organisation of examination and education, and our refusal in education to remember the wisdom of leaving some things the time and opportunity to “shift for themselves.”

It will, I think, be generally admitted that culture is something more than the knowledge of the best that has been thought and said in the world, according to the definition of Arnold, unless knowledge is used with a far wider connotation than is possible in the examination-room—something perhaps more personal and sensitive than the “enlargement” or “illuminative reason” which John Henry Newman contrasts with know-

## NATIONAL EFFICIENCY

ledge.<sup>1</sup> I shall enter into no rivalry of epigram with these great masters of style ; but I think, on cross-questioning, we should find that the cultured man is the man who is sensitive and responsive over a large field of knowledge regarded as of value by the community. On further cross-questioning we should again find it generally admitted that that sensitiveness and responsiveness are apt to suffer injury from the intellectual repression required for the examination syllabus. At every step the delicate feelers of the mind are paralysed by the suggestion : " I am wasting my time in going further ; that won't be asked." It may be held, and I should agree, that culture is as individual a thing as conscience ; that culture may be killed, that it cannot be caught, by examinations. Yet teachers who realise all this, who think examinations in their subject mischievous rather than helpful, implore the authorities to include it in every possible examination syllabus. Why ? Because, under the present *régime*, a subject that is not examined in is likely to disappear speedily from our teaching curricula. That point is sufficiently illustrated by the controversy on examinations in Greek.

We have thus seen that examinations are nowadays conducted with four distinct primary objects :—

(I) To test the efficiency of a person for the practice of his technical calling or profession.

<sup>1</sup> See Note C, p. 93, below.

## EXAMINATIONS AND

(2) To arrange candidates in an order of merit.

(3) To test the efficiency of schools.

(4) To promote general culture, an object for which many persons regard them as totally unsuited.

I am quite aware of the complexities of the issues raised by this simple classification. A recent writer, dealing with the same problem from a different point of view, has made out a list of seventeen qualities which an examination can test, and thinks his list might be extended still further. My classification deals only with groups of such qualities, and the grouping will vary with the specific object of an examination. I think it quite possible that I should find it difficult to conciliate all my critics, and most of all those who are partial to the examinations in Class (4) of my list. I am however willing, as I said before, to concede that any particular examination, and any particular examining body, may have all the virtues claimed for them; but I would ask of those concerned if it is not a fact that if we took the examination-papers set at many examinations, and inspected them in detail, we should find it an impossible task to decide, from internal evidence, with what object, or with what combination of objects, those examinations are organised. Individual examiners are not asked to bear in mind, nor is it, perhaps, always desired that they should bear in mind, the careers which the examinees propose to follow; they are asked,

## NATIONAL EFFICIENCY

not to examine with any particular aim in view, but to examine "in their subject," and this, I suppose, is what is meant by the opponents of examinations when they rail against examination for examination's sake. Thus I am inclined to think that the most ardent advocate of the examination system would be prepared to admit that the object of many examinations held nowadays is indeterminate. It can scarcely be expected, therefore, that the results will be in all cases precise.

Before we can estimate in detail the results of examinations we must say something of their methods. Examination-tests may be classified in two ways: first, according to what may be called their external character, under which heading we may divide them, as usual, into written examinations, oral examinations, and practical examinations; and, secondly, according to a character much less obvious, but more important. They may be divided into:—

(1) Knowledge-tests (or memory-tests),<sup>1</sup> which test the power of re-stating facts and arguments of a kind that may be learnt by rote—*e.g.* "Describe Joule's experiments on the mechanical equivalent of heat"; "Relate the chief events in the reign of Edward I."; "Write out the 47th proposition of the first book of Euclid"; "How is the preposition *in* used in Latin?"

(2) What may be called capacity-tests, which

<sup>1</sup> Cf. p. 52 and Note F, p. 136, below.

## EXAMINATIONS AND

test the power of doing something, of applying knowledge to some definite purpose, *e.g.* the power of making a *précis* of a written document, of writing a letter or a report on a particular subject with a particular object in view, of translating from or into a foreign language, or of conversing in that language, of solving a mathematical problem, of criticising a passage from a literary work, of writing an essay on an historical or literary subject with the aid of books in a library, of diagnosing the malady of a patient, of analysing a chemical mixture or compound ; and (perhaps, in certain ways, the highest form under the rubric) of making an original contribution to learning or science as the result of personal investigation or experiment.

In some cases the category to which a question belongs will depend on the previous preparation of the candidate. If a candidate who is asked to describe the foreign policy of Queen Elizabeth has had access only to text-books in which Elizabeth's foreign policy is treated incidentally in a more or less chronological record of events, he may show considerable capacity in separating out in his mind the threads of foreign policy from the rest, and in putting the results of his thinking into clear language ; if he has been in the hands of a teacher who has dwelt on the particular question, his reply may show memory or " knowledge " and nothing more. In subjects like political history or history of literature, it appears to me extraordinarily diffi-

## NATIONAL EFFICIENCY

cult to set questions in papers of the ordinary three-hours type which shall come clearly under the heading of capacity-tests<sup>1</sup>; whereas in many subjects the questions may be either knowledge-tests or capacity-tests, and most written examination-papers include questions of both types.

The classification which I have here set down, hinted at by Mark Pattison, and no doubt by earlier writers, but perhaps first clearly enunciated in a remarkable but too little-known work by Latham on "The Action of Examinations," published in 1877, obviously does not correspond to a logical dichotomy. The capacity-test of the kind used in examinations does imply knowledge, though the knowledge-test does not imply capacity beyond that of memorising, and an elementary (and at times doubtful) mastery of the mother-tongue. It is sometimes said that the memory-tests set at examinations are too hard; that the knowledge laboriously pumped into the brains of candidates under examination-pressure escapes rapidly when that pressure has been released; and that if we were to apply the witty definition of education as what remains when we have forgotten what we have learnt, we should find the result of examination-education to be *nil*. It would not be wise to

<sup>1</sup> Mr. M. W. Keatinge, in his *Studies in History* (A. & C. Black, 1910, chap. viii.), makes interesting proposals to meet these difficulties. Sir Napier Shaw, F.R.S., in an interesting pamphlet privately printed, suggests that our curriculum is largely determined by the form of the ordinary three-hours' paper.



## EXAMINATIONS AND

accept that conclusion without reserve. It is, at any rate, true that examiners do not by any means demand that candidates should know everything that they are asked ; the ordinary examiner at a pass examination is content if the examinee can answer satisfactorily from one-third to one-half of the questions set. When these questions are memory-tests, the indulgence is surely justified in most cases, although there are perhaps ignorances of fact of a dangerous character to which this leniency should not apply. An examiner once told me that on the ground that a certain candidate had obtained the necessary number of marks in physiology (and I shall come to the question of marks presently), he had been allowed to pass, although he firmly believed that there was air in the left ventricle of the human heart. The examination, I may say, was not a University examination. On the other hand, for some examiners printing might never have been invented, and books of reference might be as rare as they were in the Middle Ages. I shall not be accused of exaggeration when I say that a knowledge of technical terms in foreign languages is sometimes demanded by examiners which would certainly not be expected of candidates in their own language.

In dealing with capacity-tests, the question of leniency and severity takes on another aspect. A boy who can do right only five addition sums out of ten cannot add. A person who reads a ther-

## NATIONAL EFFICIENCY

mometer accurately five times and inaccurately five times cannot read a thermometer. A person who understands nine-tenths of the words in an easy passage in a foreign language, with or without the use of a dictionary, but is at sea in regard to the meaning as a whole, has not brought his knowledge of the language to a useful point. On the other hand, if a candidate can satisfy capacity-tests, if we are able to assert from the definite evidence of the examination-room, by his obtaining not 30 or 50, but, say, 80 or 90 per cent. of the maximum marks assigned to such tests,<sup>1</sup> that he can do this kind of problem in mathematics, that he can speak this foreign language so as to make himself understood on everyday topics, that he can analyse a leading article in a daily paper so as to present its gist fairly and clearly to a person who has not read it, we need not fear that these capacities will fade away a week or a fortnight after the examination. I firmly believe that examinations *can* tell us something definite about candidates. The question is, do they at the present moment? It is obvious that the answer will not be the same in regard to all examinations—that no general answer can be given to my question. But we shall understand its precise bearing if we deal a little more in detail

• <sup>1</sup> Cf. the evidence given by Prof. Gabriel Lippmann, the distinguished Professor of Physics at the Sorbonne, at the *Enquête sur l'Enseignement Secondaire*, conducted by the Chamber of Deputies in 1899, *Procès-verbaux*, vol. ii., p. 35.

## EXAMINATIONS AND

with the question of numerical marks, adopted, not in all examinations, but in most pass examinations and in most competitive examinations.

The system is familiar to every one. A certain number of marks is assigned to each question set, and the maximum obtainable for a paper is the sum of these marks. The total number of marks allotted to each candidate is the sum of the marks assigned to his several answers to those questions ; and the ordinary working hypothesis adopted is that the value of the performance of the candidate is proportional to the number of marks he thus obtains. Putting the thing in another way, we may say that the hypothesis amounts to this, that, if we arrange candidates in the order of their performance as determined by the aggregate of marks obtained, we shall be doing them no injustice ; and this order is commonly called an order of merit (though order of proficiency would perhaps be a more accurate expression).

The value of the hypothesis may be tested in two ways. We may examine the postulates on which it rests, and the results to which it leads. Take, first of all, the postulates. The hypothesis implies that we have some reasonable basis for determining the relative values of a series of questions, *i.e.* that a number of different examiners, regarded as equally competent, will attribute approximately the same relative values to the individual questions ; and it implies, again, that

## NATIONAL EFFICIENCY

the same set of examiners will attribute approximately the same values to the answers to those questions given by a particular candidate. Now, although examinations are carried out on such a gigantic scale, we have very few published data to test the validity of these postulates. In a statistical investigation by Professor Edgeworth, it appeared that there might be very great discrepancies in the marks assigned by different examiners to the answers of a set of candidates.<sup>1</sup> If this is so, how can the same standard be maintained? The result is what we might expect in the case of examinations of which the object is indeterminate, for how can we estimate the value of anything unless we know what is its precise object? If a candidate is allowed to pass by obtaining from 30 to 50 per cent. of marks on memory-tests and capacity-tests mixed in variable proportions, what does a "pass" mean? What may we say that a man who just passes an examination can do? Is not the meaning of a "pass" in itself indeterminate? Does not the standard necessarily vary from candidate to candidate as well as from year to year?

If we turn again to the results of examinations carried on on a large scale, there are indications that the results of this working hypothesis are not more satisfactory than its postulates. Examiners themselves are often heard to say, "This man is better than his marks, this man is worse

<sup>1</sup> See Note E, p. 118, below.

## EXAMINATIONS AND

than his marks,"<sup>1</sup> although perfect freedom may be left to them to assign marks so as to correspond to resultant merit. And if we inspect published statistics of examinations for which a large number of candidates present themselves, and in which it would be reasonable to expect that approximately the same number would attain a pass standard in consecutive years, they show that there may be the most extraordinary variations in the percentage of passes in the same subject. Those variations, as I think any one acquainted with the inner working of our examination system will agree with me in saying, cannot be regarded as due to the carelessness of individual examiners; for, in the first place, examiners take a great deal of pains with their work, and, in the second place, carelessness may tell as much in favour of a particular candidate as against him. I do not think that it is at present possible, without further scientific investigation on a considerable scale, to state what are the precise facts of the case, or the precise causes of these variations. They may be due to defects in our system of marking; they may be due to real variations in the quality of candidates, of which the causes are unknown and which it would be important to discover.

I want to strengthen my plea for investigation by quoting a witness whom I might have expected to find against me. If there is one subject in which accurate marking and community of

<sup>1</sup> Cf. p. 60, below.

## NATIONAL EFFICIENCY

standard would appear to be most easily obtainable, it is mathematics; and it seems probable that the whole system of marking by numbers is derived from the marking of mathematical papers. Nevertheless, Professor Bryan, a distinguished mathematician and a high authority on the subject, regards some of the troubles to which I have referred as due to "fetish-worship" of the maximum mark, and suggests that we should assign certain marks to correct answers of average quality, and then mark up to any maximum we please, to reward exceptional proficiency shown on any part of a paper. In this way, he thinks, candidates would no longer feel the inducement to "scramble for marks by scribbling scrappy attempts." Other authorities have suggested different ways of dealing with these difficulties, but it is safe to say that there is no agreement, and that different examiners employ very different methods. The  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ , or Oxford system is another form of numerical marking in which the total number of marks allotted is low,<sup>1</sup> and in which, if I am rightly informed, the examiners as a rule attempt to obtain their resultant impression of a paper as a whole directly, instead of by the addition of marks assigned to individual

<sup>1</sup> Articles on "Examinations," by Prof. G. H. Bryan, Sc.D., F.R.S., published in the *School World* for May and June, 1910.

<sup>2</sup> By adding the signs + or - to the Greek letters, the number of symbols used is, as a rule, increased to ten or more. Cf. p. 62, below.

## EXAMINATIONS AND

questions. Here, again, it would require investigation to decide whether this system is or is not more trustworthy than the other. In any case, in determining a man's pass or class, you do not escape entirely from the difficulties involved in the addition of marks. Moreover, the Greek letter system appears, so far as I can see, to be inapplicable to a competitive examination on a large scale in which candidates present themselves in a number of subjects.<sup>1</sup>

The assignment of marks in practical examinations is probably an easier matter than in written, for the reason that most practical examinations, though not all, are capacity-tests rather than memory-tests, and it is easier to say whether a candidate can or cannot do what is required of him.

Of all kinds of examinations, that which may be the most valuable, and that which presents the most difficulties, is the *viva-voce* examination. Largely used abroad, it is used comparatively little in this country, because our boys and girls are not taught to speak and to think at the same time at the age when children speak confidently and easily; and hence, in the presence of the examiner, the candidate is often paralysed by nervousness, for which it is difficult, or impossible, to make the proper allowance. He may be reduced to silence or nonsense. A candidate in *Materia Medica* in a northern University, who was unable, owing to

<sup>1</sup> See however p. 62, below.

## NATIONAL EFFICIENCY

examination fright, to distinguish the taste or the smell of the sample of cod-liver oil that he held in his hand, on being asked, "Where does cod-liver oil come from?" replied, "The whale."

Nevertheless, the *viva-voce* examination has the immense advantage of allowing a reasonable examiner to ask a candidate not paralysed by nerves what portion of a subject he has specially studied, and to cross-examine him considerably in that. The negative evidence that a candidate says nothing in regard to a particular printed question does not carry one very far. The frank "I don't know" of a candidate, his power of distinguishing what he does know from what he does not, will, with an examiner capable of judging men, count very greatly in favour of a candidate whose knowledge of the parts of the subjects that he thinks he knows is sound. But a *viva-voce* examination, if it is to be valuable, must not be hurried, and hence it is an expensive form of examination. It has again the defects of its qualities. The idiosyncrasies of examiner and candidate may come into conflict in regard to matters which have nothing to do with the particular examination-test. To combat this defect, it is generally recognised that *viva-voce* examinations should, if practicable, be conducted by a board of examiners rather than by a single examiner.

We have in the *viva-voce* examination a sense of reality that is apt to become somewhat thin in the written examination. If an examiner knows that



## EXAMINATIONS AND

he has to decide, face to face with a candidate, not whether he attains a particular standard in a particular subject, but whether he has the knowledge and capacity to do a particular job or pursue a particular career, he will feel much more able to come to a decision than after he has merely seen the candidate's written papers. Would any sensible person, in business, in war, or in administration, free to choose his own subordinates in carrying out an important piece of work, dream of doing so without an interview, if he could help it? On the whole, it seems to me that our examination system would be more efficient if larger use were made of the oral method than is made at present. But there are administrative difficulties in conducting examinations by that method not to be under-estimated.

Before leaving the question of examination methods to inquire further into the effects of examinations on the candidates, I want to throw out one suggestion. I believe that we should effect a quiet revolution in pass-examinations, but a revolution, if we would systematically ask ourselves in regard to each examination, What does it prove that all successful candidates can do? What are they certainly fit to do? Such a question would put capacity-tests and memory-tests, I believe, in their right places, and the examination would become once more, as in its origin, what the mathematicians call a directed quantity.

And now for the effect of examinations on candi-

## NATIONAL EFFICIENCY

dates. If we take a large cold thermometer, as accurate an instrument as we please, and plunge it into a relatively small quantity of warm water, the reading of the thermometer will not tell us directly what the temperature of the water was before we tried to measure it. In the same way, even the most admirably designed and organised examination will generally exercise a very considerable disturbing influence on the previous education of the candidate that it is designed to test, and on the candidate himself during the examination. In the case of the thermometer, we can correct our original reading by making the necessary allowance ; in the case of the examination candidate, what correction can we possibly make ?

How, if we cannot calculate them precisely, shall we reduce the bad influences and magnify the good influences of examinations as far as possible ? It is a grave problem, and we shall find considerable disagreement in determining what features of examinations belong to one category, and what to the other. Let us, for the moment, consider the influence of the examination on the candidate while it is being actually carried on. There can be no doubt that examinations prolonged over several days, for from four to six hours a day or more, involve a considerable physical strain on the candidates, and the sample of knowledge and capacity which the examination affords can hardly be regarded as an average

## EXAMINATIONS AND

sample. With some candidates the excitement and strain make them do better than they otherwise would do ; with others the reverse is the case. Ought we to try to reduce that strain ? A good many, I fear, would not assent. It has been said, and not infrequently said, that if a candidate fails under the strain, he is lacking in some of those qualities which go to make a successful man, and that though failure in the particular examination does not mean (as the public might, and often do, think) that he has not the intellectual attainments of those who pass, yet that no harm is done. It seems probable—one cannot go further than that—that failure in examinations owing to temporary ill-health takes place more frequently in the case of women candidates than men, and the question of women candidates certainly needs separate investigation ; but in the absence of adequate statistical and medical inquiry, the whole hygiene of the examination-room remains obscure. Much more important, in the long run, than the immediate influence of the examination-room on the candidate is the influence of his preparation for the examination during the long years of school and university life. That is a subject too complex and important for me to do more than mention ; I would only suggest two things—first, that in some subjects and with some candidates, the influence of examinations may be, and often is, good and not bad ; and secondly, that instead of encouraging by means of examination subjects

## NATIONAL EFFICIENCY

ill-suited for examination-tests, we might try to protect them from those tests. I do not say that it is easy, but I think it not impossible.

Let me take the subject of history, which, of all subjects that come under the heading of "Culture" subjects rather than bread-and-butter subjects, is the most important to the ordinary citizen. Would it not be possible to reduce school examinations in history to a minimum by allowing those who had attended a history class in a secondary school, to the satisfaction of a duly qualified teacher, to be exempted from any kind of school-leaving examination in history? Of course, it will be said that the standards of different teachers would be different, but are we not inclined to sacrifice on the altar of equality the very efficiency of the teaching that examinations are intended to test? I remember, in a French Lycée, hearing a lecture on Napoleon's *coup d'état* of the 18th of Brumaire given to a class of science boys who were not to be examined in the subject. The lecture was full of warmth and colour, every boy was interested, every boy was listening; and I contrast that with the painful dictation of slow sentences, reproducible in the examination-room, that I have heard in the history class-rooms of some English secondary schools. I do not mean that all French teaching of history is good, and that all English teaching of history is indifferent. I believe there is no subject in which more rapid

## EXAMINATIONS AND

progress is being made in this country. Might not the examining bodies help the history teachers in other and better ways than by inserting history in all their syllabuses, as the teachers seem to desire?

In dealing with this matter, I would suggest that it is as impossible to make certain that every boy or girl who leaves school has an adequate knowledge of history, has a real beginning of historical culture, as it would be to make sure that every boy or girl who leaves school has musical culture or artistic culture. Supposing you do make sure by examinations that the school-leaving certificate implies a certain knowledge of the facts of history, what then? A boy coming out of the examination-room once said to me: "I am never quite sure whether the date of Magna Carta was 1215 and Waterloo 1815, or whether it was the other way about, but I think I have got it right this time." A boy of that kind, even with the more difficult questions set nowadays, might absorb history, but he would never assimilate it, and instead of trying to force him to be cultured in a subject where culture is impossible for him, would it not be much better to recognise his insensitiveness on the historical side, and push him in some other direction? What I have said of history applies even more strongly to literature. If we could only make up our minds to require real efficiency in one or two subjects, and give our boys and girls the opportunity of learning other

## NATIONAL EFFICIENCY

subjects, without forcing them mechanically to do so, would not our education become much more effective in the end? Are we not led by the will-o'-the-wisp of general culture into a mire of general inefficiency? I asked a friend of mine who is an experienced examiner, what it would be safe to assert of the powers of a student who could just succeed in obtaining pass marks at a pass examination in several subjects. His reply was: "I think you could say that he could pass the examination at the time that he passed it." I went further. "Would it be safe to say that he had any useful knowledge of the subjects of the examination—that he would be fit for any career?" The answer was clear: "No, you could say none of these things. When you have said that he has passed the examination, you have said all that can be said."

If we turn from pass to competitive examinations, we find that they involve all the same problems, and fresh ones to boot. Here it is well recognised that difficulty may be used as the means, not of testing proficiency in subjects that will be required in later life, but to distinguish different candidates in regard to what is called their "general ability." The great choice of subjects permitted by the regulations for the Open Competitive Examination for the Home and Indian Civil Services sufficiently illustrates this point. That particular examination is one of the greatest importance. Is it really fulfilling its pur-

## EXAMINATIONS AND

pose in the best way possible? Professor E. G. Browne, of Cambridge, in giving evidence before Lord Reay's Committee on Oriental Studies (Minutes of Evidence, page 62), said that the men selected by the method devised by Lord Cromer, in connection with the Egyptian Services, were more apt in the learning of Oriental languages than the men selected by open competition for the Indian Civil Service.<sup>1</sup> I believe Lord Cromer intends to say something, later, on the methods that he used. I do not say that Professor Browne's isolated observation proves anything conclusive, but it is significant, and points to inquiry.

The tremendous severity of the Civil Service examinations has been regarded as imposing a sufficient test of the character and of the physical strength required for arduous public service. But do we know at the present moment whether the game is worth the candle? A letter from a candidate, who wrote over the somewhat unfortunate signature "Presque passé," in the *Times* of last summer, reminds one uncomfortably of the sad picture of overworked Indian examination students drawn about the same time by Mr. [now Sir] Valentine Chirol in his articles on Indian

<sup>1</sup> Minutes of evidence of Treasury Committee on the Organisation of Oriental Studies in London (Cd. 4561), 1909, Questions 1740-50. Prof. Browne attributed the difference to the prominence given to mathematics and science in the competitive examination.

## NATIONAL EFFICIENCY

Unrest.<sup>1</sup> Is it not possible to test "general ability," and to separate the ablest candidates by methods involving less strain both on the successful and the unsuccessful candidates? The experimental psychologists and statisticians, working on lines suggested by the late Sir Francis Galton, hold out hopes in this direction, but one must confess that they are still somewhat remote. Would it be possible, without reintroducing the unspeakable evils of jobbery, to follow the lines laid down by Lord Cromer in the Egyptian Civil Service, and by Lord Selborne in choosing candidates for the Navy? These are questions again for inquiry.<sup>2</sup>

<sup>1</sup> Reprinted as a volume, *Indian Unrest*, 1910, see pp. 217-18.

<sup>2</sup> Referring to examinations for the Indian Civil Service, Army, and Home Civil Service Clerkships of the Second Order, Prof. Edgeworth writes (*Journal of the Royal Statistical Society*, vol. liii., p. 653):—"I find the element of chance in these public examinations to be such, that only a fraction—from a third to two-thirds—of the successful candidates can be regarded as quite safe, above the danger of coming out unsuccessful if a different set of equally competent examiners had happened to be appointed." Without being an expert in statistics, I venture to think that the basis of Prof. Edgeworth's calculations is too slender, and now too much out of date, for us to draw general conclusions from them; the uncertainty of the examinations in question may be to-day much greater or much less than that shown by those calculations. The serious matter is that we are entirely in the dark as to the actual situation. [I reprint this footnote as being an integral part of the original paper. For a fuller account of Prof. Edgeworth's work, see Note E, p. 102, below.—P. J. H.]



# EXAMINATIONS AND

## *Conclusion*

How should such inquiry be carried out? I do not think that any other body than a Royal Commission would carry with it the necessary weight. I believe the Commission should be a small one, presided over by a statesman with great experience of affairs, and that there should be no attempt to achieve the impossible by including in it representatives of all parties concerned. If, following the example set in Professor Sadler's Moral Education inquiry,<sup>1</sup> and in the recent report of the Consultative Committee of the Board of Education on Evening Schools, the evidence taken were published, not in the unreadable and unwieldy form of question and answer, but in the form of statements on particular topics summarised by the witnesses from shorthand notes, after their examination by the Commission, every person and every interest would have a fair say; and the members of the Commission themselves would not have any undue advantage in giving their own summary of the evidence, and their own conclusions. But such a Commission, to do all that it should do, would need to be far more than a Commission of Inquiry in the ordinary sense; it would need to be a Commission of Investigation. It should have the power to appoint

<sup>1</sup> *Moral Instruction Training in Schools: an International Inquiry*, ed. by M. E. Sadler (Longmans), 1908.

## NATIONAL EFFICIENCY

and to pay committees of experts in different subjects, who would carry out investigation under its direction. Take the questions of standard and of marking. Those questions it is difficult to investigate adequately without having a large number of candidates' answers lithographed or printed and sent to different examiners to be marked, first without any directions at all, and secondly with directions relating to the stress to be laid on particular points regarded as of especial importance. Nor would the expert in his subject alone be necessary; the assistance of the statistician would have to be called in at each step. I believe that in this way we might be able to make an examination standard that would mean something very different from what it means now, that would be reasonably, though not rigidly, constant from year to year, and that would substitute the fairness of considered judgment for the unfair impartiality of chance, in the case of far more candidates near the borderline. That would be one of the matters for the Commission to investigate. If it were to appoint sub-committees in a half-dozen, or a dozen, subjects of examination, the results of the inquiry would probably throw sufficient light on the subjects which it was not possible to touch.

A second question for investigation is the hygiene of the examination-room. How far are we making fair demands on our students in the examinations? How far do they exceed, from

## EXAMINATIONS AND

the point of view of strain, what it is reasonable to demand ?

A third point for inquiry, and that the most difficult, is the influence of examinations, good and bad, on education as a whole. I would only suggest here that different classes of candidates and different subjects require different treatment, and that it should be the object of the Commission to find some way of protecting from examination those school subjects which are unsuited to them, and in which examinations dig up the roots of the plants which we wish to grow.

Fourthly, the Commission would inquire how far our different sets of competitive examinations select the persons most fitted for the posts which they are intended to occupy.

In this connection, the Commission would, no doubt, consider the question of associating as examiners with the experts in their particular subjects administrators with experience of the services for which candidates were being chosen.

I have selected four large topics for inquiry : (1) the question of standards and marking, with which is closely connected the question of the precise object of the examination concerned ; (2) the hygiene and psychology of examinations ; (3) the influences of examinations, good and bad, on previous training ; (4) competitive examinations for the Services. But these topics would subdivide, and others would suggest themselves for investigation. A Commission with a suitable

## NATIONAL EFFICIENCY

reference would find no lack of useful work to do. I suggest as a reference: "To investigate and report upon the methods and efficiency for their purpose of examinations carried on by Government Departments and other public bodies in the United Kingdom; to inquire into the influences of examinations on the previous education of candidates; and to suggest such changes as may seem desirable."

The investigations or such a Commission would be of immense assistance both to the teaching and to the examining bodies of this country. I have tried to obtain figures showing the magnitude of our examination operations. Those figures are still far from complete, but I have official statistics, supplied to me by the authorities concerned, which show that in the year ending September 30, 1910, out of 146,741 candidates at public examinations in the United Kingdom, 71,114 failed. I have little doubt that when my figures are complete—and I hope I may be permitted to publish them later—the total entries will approach 200,000 and the failures 100,000.<sup>1</sup> Does there not seem to be something very wrong in the relations of our examinations with our education for nearly half the persons to be judged unfit to pass the test for which they present themselves? How far is this great percentage of failures due to a low average

<sup>1</sup> Several bodies which were good enough to furnish me with statistics made it a condition that the details should not be published.

## EXAMINATIONS AND

of intelligence or industry, how far to the fault of the teachers, or the fault of the examiners, or to want of co-ordination between teacher and examiner? Conscientiously conducted in their details, have not our examinations been largely at the mercy of hazard in regard to principles?

Lord Rosebery, speaking in 1907 as Chancellor of the University of London,<sup>1</sup> urged strongly that examiners needed training no less than teachers. He also suggested the need for investigation. For such teaching as Lord Rosebery desires I doubt if any common basis of doctrine at present exists. It must be created.

There is one serious obstacle to change. The examiner and the organiser of examinations are often so weary of their ungrateful tasks that for them the day's work is enough; they become hardened, sceptical of improvement, reconciled to evils that they have come to regard as unavoidable evils. But the issues are too grave to be overlooked, and the investigation and reform that I have ventured to demand are a matter of national necessity.

### SPEECH BY THE LATE EARL OF CROMER, O.M., G.C.B.

Lord Cromer, speaking from the Chair, said:—  
I am sure that you will all agree with me when I say that we owe a deep debt of gratitude to Mr.

<sup>1</sup> At the opening of the London Day Training College; see the *Times* of November 4th, 1907.

## NATIONAL EFFICIENCY

Hartog for the very interesting paper to which we have just listened. It is a paper which stimulates thought, and, let me add, there are few subjects more deserving of thought at present than that to which Mr. Hartog has drawn our attention. We seem to be about to enter a period during which, whether for good or evil, the State, either in its imperial or municipal capacity, is likely to develop great activity in the performance of a variety of new functions. How are the large additional number of agents whose services will thus be required to be recruited? The results which will be obtained by the new departure now being made will, in a great measure, depend upon the answer which is given to this question. It is perhaps an exaggeration to say with Pope that "Whate'er is best administered is best," but it is certainly true that a good system may be marred by mal-administration, whilst the defects of a faulty system may be greatly mitigated if the administration is efficient.

There are, so far as I know, only three methods under which public functionaries can be appointed. These are election, selection or nomination—the two words are, for the purposes of my present argument, synonymous—and competition, either open or restricted.

I need say but little of the elective method. Its advantages are mainly political rather than administrative. An eminent French economist, M. Leroy Beaulieu, has said: "According to

## EXAMINATIONS AND

democratic theory, popular election will always secure the most capable individuals ; but this is a curiously mistaken idea, and is quite unsupported by experience."

The Poor Law Commission, over which Lord George Hamilton recently presided, took the same view. "The method of popular election," the Commissioners reported, "is not one which is adapted to get the best administration. The Commission believe that there is a better chance of good administration if the members are appointed than if it is left to the chance of election." I will not, however, dwell any further on this branch of the subject, as it touches the fringe of political issues, which I am anxious to avoid. I will only, therefore, say in general terms that one, and perhaps not the least important of the questions which the democracy of this country will have to face in the near future, is how to combine administrative efficiency with an adequate amount of popular control. I have often, in the course of my Eastern experience, had to deal with a very similar problem, and I generally found that the question practically resolved itself into this — How much efficiency is it justifiable to sacrifice on the altar of self-government ? I came to the conclusion, under the special circumstances with which I had to deal, that it was impossible to generalise on this subject, and that each separate issue, as it arose, had to be decided on its own merits.

## NATIONAL EFFICIENCY

Turning to the other alternatives, viz., selection or competition, it may readily be admitted that competition, which is the prevailing system now, has its defects, but before condemning it, it is worth while to glance back at the system which it supplanted. I am one of the band, not now very numerous, who can speak with some personal experience of the pre-competition days, and, as illustrations of the working of the then existing system, I may perhaps be pardoned if I relate a few biographical details. Some fifty-six years ago, being then a boy of fourteen, I was admitted into the Royal Military Academy at Woolwich, with a view to eventually becoming an officer in the Artillery or Engineers. At that time there was a sort of sham entrance-examination, but I never heard of any boy who had been nominated by the Master-General of the Ordnance being rejected. Lord Hardinge, the grandfather of the present Viceroy of India, who was a friend of my family, gave me my nomination. On presenting myself at Woolwich for medical examination, I was very rightly rejected for short sight. I returned to London and told my mother, who was my only surviving parent. She acted with promptitude. She instantly rang the bell, ordered her carriage, and went to the Horse Guards to see Lord Raglan, who had succeeded to Lord Hardinge's place, and who was another friend of my family. The result was that next day I returned to Woolwich with a letter addressed



## EXAMINATIONS AND

by Lord Raglan to the medical officer, asking him to "reconsider the matter." I was, of course, admitted. Exactly the same thing happened at the same time to another lad who was at first rejected on the ground that he had a serious impediment in his speech, but whose case was subsequently, under pressure, "reconsidered." This sort of thing could not happen now. The practical working of the system may, however, best be illustrated by an anecdote which is related of Lord Melbourne. A friend of his who occupied a high position and who disposed of a good deal of patronage, said to him: "I do not in the least mind confessing that if I had to deal with two candidates, one of whom was the son of a friend or relation of mine, and the other a stranger, I should, *ceteris paribus*, give the appointment to the son of my friend or relation." To which Lord Melbourne drily replied, "So should I, but *ceteris paribus* be d——d." In fact, at that time, the principle laid down by George III had only been but slightly modified. That monarch is believed to have said that any one was fit to occupy any place he could manage to get.

Well, ladies and gentlemen, as a result, partly of the attention called to military matters by reason of the administrative breakdown during the Crimean War, and partly of the extraordinary call for officers occasioned by the Indian Mutiny, the nomination system, in so far as appointments to the Artillery and Engineers was concerned,

## NATIONAL EFFICIENCY

was swept away, and the competitive system introduced. It started under rather bad auspices. I remember that the first notice issued by the War Office was rather unfortunately worded. They called for "persons" to compete. The result was that all the young men who responded to this call were instantly dubbed "persons," and commanding officers showed a marked preference for those who had entered under the old system, and who had enjoyed the more exalted title of "gentleman cadets." I was at that time an ardent young Liberal, as I claim to be an old Liberal now—although my right to that designation may perhaps not be recognised by mere party politicians—and I was warmly in favour of reform. As a matter of fact, the gloomy predictions of the anti-reformers were falsified by the event. The social class from which the officers who entered the Army were drawn was just the same after as it had been before the change of systems, whilst there can be no doubt that a distinct improvement was effected, both in intellectual standard and professional capacity.

I mention these points mainly in order to show that the competitive system, whatever may be its defects, is greatly superior to that which it superseded, and in order to warn the advocates of reform that they must beware lest a door be opened which may lead to a revival, perhaps in a somewhat different form, of the abuses of the past.

## EXAMINATIONS AND

And now let me make a few remarks on the competitive system considered on its own merits. I think the best thing I can do is to deal with facts which have come within my personal experience ; and, as you are probably aware, my personal experience has been gained more in the outlying portions of the Empire than at home. I have seen ~~very~~ good deal of the work performed by the young men who are turned out by our universities and military colleges. I have met them on the banks of the Ganges, the Indus, the Irrawaddy, and the Nile. And I have also had some rather exceptional opportunities of comparing their aptitudes with those displayed by their Continental rivals and competitors. From one point of view the comparison is disappointing. I have often had it said to me by heads of departments in Egypt, themselves Englishmen, that the young Frenchman or German came to them crammed to the muzzle with an amount of preliminary knowledge greater than that possessed by the young Englishman. The explanation is, I think, that until recent years sufficient attention has not been paid in this country to technical education ; that is unquestionably a defect, more especially at a time when there is a tendency to specialise every branch of learning, but it is a defect which I hope and believe is in process of being remedied. But if we look at education not merely as a means for storing the mind with knowledge, but also as an instrument for developing the character, the

## NATIONAL EFFICIENCY

situations are reversed. The youth in whom the need of individual effort and the sense of personal responsibility have been nurtured in one of our public schools or colleges, becomes almost unconsciously *capax imperii*. He is no automaton. In the free atmosphere in which his boyhood and youth are passed, he learns a number of lessons which stand him in good stead in after life as a member of an imperial race. More especially, he learns not only to think, but, which is much more important, to act for himself—in a word, to govern, and to govern generally with firmness, justice, and wisdom. From this point of view our system of making appointments, which is generally, though not exclusively, the result of competition, may, I think, in spite of some defects, be said generally to have produced satisfactory results. A very high authority on Indian affairs, Mr. Valentine Chirol, recently wrote: "To the Indian civilian belongs the credit of almost every measure passed during the last fifty years for the benefit of the Indian masses, and passed frequently in the teeth of vehement opposition from the Indian politician." I think you will agree with me that this is very high praise. The system has thus turned out a number of young men who, as the agents of a nation whose very existence depends upon the execution of a sane imperial policy, are not only of incomparable merit, but, I do not hesitate to add, are also the envy and admiration of the whole civilised world. What-

## EXAMINATIONS AND

ever changes may in the future be made in our educational system, I most earnestly trust that no steps will be taken which may tend, in any degree, to weaken that sense of independence in thought and action which should constitute the basis of character to the individuals who are the units of any self-governing community.

In spite, however, of this success, I think we may well ask ourselves, after the experience of half a century of the working of the competitive system, whether it is not capable of improvement. The question is certainly worthy of consideration. Remember that the principle of selection, if only it can be properly carried out, possesses merits superior to those of competition. The former may, or ought, to result in the creation of leaders of men. It should enable us to acquire for the public service the aristocracy, in the proper and original sense of the term ; that is to say, not the aristocracy of birth, but that of intellect and of character. The latter tends rather to produce a dull level of mediocrity. Most educated Englishmen have recently been reading the history, written by a very distinguished contemporary statesman, of one of his most illustrious predecessors—I mean Lord Rosebery's history of the early life of Chatham. Now, what was it that enabled Lord Chatham to revive British spirit when it had sunk to a very low ebb, to march to victory on the banks of the St. Lawrence and the

## NATIONAL EFFICIENCY

Ganges, and to found the British Empire: I say that it was the adoption of the principle of selection, sparingly, indeed, but very skilfully employed. I notice with pleasure that of late years there has been a tendency, notably in the military, naval, and diplomatic services, to adopt this principle in dealing with all the later stages of the careers of public servants more thoroughly than was formerly the case. I trust that this movement, far from being arrested, will be pushed still further. I could name numberless instances; especially in the military service, in which the dead weight of rigid regulations, based generally on the claims of seniority, have acted detrimentally to the public interests.

The case of first appointments presents naturally far greater difficulties, for, *ex hypothesi*, there can in these cases be no record of past services on which to rely, and moreover, here, and indeed in any application of the principle of selection, we are at once met with the awkward question, *Quis custodiet ipsos custodes?* Who is to select the selectors?

In connection with this subject, there is one point in respect of which I cannot help thinking that a good deal of misapprehension exists. It is not unfrequently thought that Ministers and others in whose hands the bestowal of patronage is vested, would rather welcome an accession to their powers in the matter of making first appointments. I believe this view to be wholly erroneous.

## EXAMINATIONS AND

My experience has led me to the conclusion that those who exercise patronage are also those who most of all delight in being fettered by regulations which relieve them of a very irksome and onerous duty, in the exercise of which they are certain to incur a great deal of misrepresentation and abuse, and to receive very little gratitude. It is, in fact, true that the wider is the discretion left to a Minister the more fully will he be exposed to the full blast of social, and also—which is perhaps even more hurtful—of parliamentary pressure, to make his selections on grounds other than those of merit. Moreover, unless he be an exceptionally strong man, it is well-nigh certain that he will at times succumb to this pressure. More than this, it has to be remembered that neither the action of Parliament itself, in its collective capacity, nor the power exercised by the Press, however honest in intention, affords any adequate guarantees either against deliberate jobbing, or what is perhaps a greater danger, against unintentional negligence. The truth is that, save in a very few exceptional cases, the number of persons who can speak with impartiality and full knowledge of the merits of a candidate, whether for promotion, and still more for first appointment, is very limited, and, moreover, it is not unfrequently difficult to obtain a frank expression of their opinions. I do not suppose, therefore, that any one, however fully he may recognise the defects of the competitive system,

## NATIONAL EFFICIENCY

would be prepared to advocate a return to the principle of uncontrolled selection.

Let me again fall back on my personal experience, and explain how, in one special instance, to which Mr. Hartog has alluded, I solved the difficulty myself. Some few years ago it became necessary to create a Soudanese Civil Service. In the first instance the appointments were practically made by myself. I found it a very difficult task. I used to receive hundreds of letters from the friends or relations of candidates. Of these about 90 per cent. found their way into the waste-paper basket. I took an enormous deal of trouble to select the best men. Indeed, I may remark incidentally that whatever success has attended the administration of Egypt during the last thirty years has been mainly due to the great care which was taken both in selecting and in promoting officials. Generally speaking, the method adopted in the first instance in the Soudan for making first appointments worked fairly well. Nevertheless, I felt that, apart from the great trouble and responsibility it imposed on myself, the system, or perhaps I should rather say, the absence of system, was very unsatisfactory. In the first place, I never could feel sure that the merits of many young men perhaps more deserving than those who were actually appointed had not been overlooked. In the second place, although I felt quite certain that I should not job, I felt almost equally certain that,



## EXAMINATIONS AND

sooner or later, I should be accused of jobbing, and the accusation would in itself have done harm, inasmuch as it would have shaken public confidence. On the other hand, I was not prepared to resort to open competition, as I was convinced that many of the qualities most required for a Soudanese civil servant could not be tested by a competitive examination. Eventually a system was adopted under which the duty of selecting candidates, who are all taken from the universities, was, under conditions which I need not describe in detail, conferred on a board of highly qualified officials, who were left quite free to act as they thought best in the public interests. The system produced good results, but I do not at all say that on that account it is suitable for general adoption in this country. There are alternatives on which I need not dwell. My present object is merely to draw attention to the subject, and to indicate the advisability of considering whether it is not possible to adopt measures which will on the one hand admit of some independent control being exercised over State authorities, and more especially local authorities, in cases where the competitive system, properly so called, is not applied, and which will, on the other hand, tend to remedy, or at all events to mitigate, the acknowledged defects in the competitive system itself. The test supplied by that system, though in some respects valuable, is purely intellectual, and

## NATIONAL EFFICIENCY

indeed, as at present applied, it may be doubted, more especially after the evidence adduced by Mr. Hartog, whether, even as an intellectual test, it always succeeds in attaining its object. Time does not permit me to dwell on the details ; I will therefore only say that, in laying down the conditions under which a competitive examination is to be conducted, the most important points to be considered are, first, how are the examiners to be appointed, and, secondly, on what principles should they proceed in awarding marks.

For the rest, I should like to point out that the Order in Council of 1870 contemplates the possibility of a departure from strict rules in certain cases, but such a departure can only be made with the consent of the Civil Service Commissioners. It is in the development of this idea that, for my own part, I should be inclined to look for reform. In any case, I hold a strong opinion that if anything is to be done in this direction, Parliamentary politicians should be rigorously excluded from participation in any Boards or Commissions which may have the disposal of patronage.

Ladies and gentlemen, I fear I have detained you too long. I trust, however, that this discussion will lead to the matter on which Mr. Hartog has dwelt being considered by others more competent than myself to deal with them.

The idea of appointing a Royal Commission to

## EXAMINATIONS

examine into the whole subject is certainly one which commends itself to me.

[*Note*.—The paper and speech printed above, together with the discussion which followed, were published in the *Journal of the Royal Society of Arts* for February 3rd and 10th, 1911. In the course of the discussion the proposals for an investigation of the examination-system were supported by Sir Henry Miers, F.R.S., Professor John Adams, and Professor Arthur Schuster, F.R.S.; they were also warmly supported by Professor M. E. Sadler in a letter printed in the *Journal* for February 3rd. I have made a few slight corrections in the text of the paper, and added one or two footnotes.]

## II

### THE THEORY OF EXAMINATIONS

#### I. DEFINITION

1. For the purposes of this article the word "examination" (except where otherwise indicated) is used to mean "a systematic test of knowledge, or of either special or general capacity or fitness, carried out under the authority of some public body." The argument, for the sake of brevity, mainly refers to examinations as carried out in the United Kingdom.

#### II. ORIGIN AND OBJECTS OF EXAMINATIONS

2. Our modern examinations are based on mediæval examinations designed with a view to testing whether a candidate had the necessary qualifications to practise a craft or profession. They have since been applied to the following new purposes: (1) to arrange candidates in order of merit (competitive examinations); (2) to test proficiency in a subject, or in one or more subjects, and, following on this, to test "general culture"; (3) to test the efficiency of schools.

## THEORY OF EXAMINATIONS.

3. The distinction between professional examinations on the one hand, in which the examiners are guided by a definite and definable (though often complex and unanalysed) object in view, and non-professional examinations on the other, in which there is no such object in view, serving as a guide, is of a fundamental character. Most of the anomalies of our examinations are found in non-professional examinations, or in those parts of professional examinations based on the model of non-professional examinations.

4. Intermediate between professional and non-professional examinations are examinations of which the purpose is definable, but of which the definition of purpose conveys widely differing ideas to different minds. The chief examinations in this category are (i) "School-leaving examinations," (ii) professional entrance examinations, and (iii) University entrance examinations (Matriculation Examinations). The first are supposed to show evidence of a good general education, and to imply fitness to enter a business or professional career, without specific limitations; the second, to imply fitness to enter on a course of study for a particular professional career (*e.g.* that of doctor or solicitor); the third to imply fitness to pursue a University career. Frequently one and the same examination is used for all three purposes.

5. The examination system has acquired an immense importance in all civilised States, and

## THEORY OF EXAMINATIONS

has become an essential factor in the social organism, since the selection of individuals to fulfil various functions in the State, as well as for the great categories of social occupation, is largely determined by examinations. It has been calculated (*Journal of Royal Society of Arts*, 1911, p. 1132) that there were in the United Kingdom in 1910 over 300,000 examination candidates in elementary and secondary subjects only, apart from University and professional examinations, and the relative numbers of candidates in the United States, France, Germany and other large countries are probably not less.

6. Nevertheless, there are few questions relating to examinations which can be regarded as uncontroversial, owing probably to two reasons: (1) the uncertainty of social and educational ideals at this time of unrest (even before the War); and (2) the fact that examination tests have never been submitted to any scientific investigation, and that we are largely in the dark as to their validity for the purpose for which they are employed, and sometimes even as to the purpose itself of a given examination and the meaning of the certificates conferred in connection therewith.

### III. CLASSIFICATION OF EXAMINATIONS

7. The kinds of test used in examinations may be classified for different purposes in a number

## THEORY OF EXAMINATIONS

of ways (which necessarily overlap to some extent), as follows :

(a) According to the method by which answers to the examination-questions are furnished by the candidates: (i) written, (ii) *viva-voce*, (iii) practical (including clinical). Examinations may combine these tests in various ways.

(b) According to the period occupied by the examination work, which may (i) be relatively short and limited by the examining body, extending over a period of from a few hours to a few days, or (ii) extend over a long period, generally without any maximum limit. Examinations of course-work done under the supervision of teachers, and of theses and dissertations, come into this second category.

(c) According as they are conducted (i) by persons not appointed to examine as teachers, *i.e.* external examiners—it is to be noted that in some, though not in all, external examinations the examiners may and do include persons who have taught some but not all of the candidates—or (ii) by teachers of the candidates, or (iii) by a combination of categories (i) and (ii), a frequent combination in University examinations in the United Kingdom.

(d) According as they are (i) knowledge (or memory) tests, or (ii) capacity-tests, or (iii) a combination of the two. A knowledge-test shows the power to reproduce in a more or less modified form information acquired by the can-

## THEORY OF EXAMINATIONS

didate ; a capacity-test shows the power to do something other than to reproduce, e.g. to speak a foreign language, to read a thermometer, to auscultate a patient.

(e) There are also psychological tests, conducted by experimental methods.

(f) Tests of medical fitness.

### IV. CRITERIA OF TESTS

8. The word *examen* means the pointer of a balance ; the analogy of the testing of the physical properties of a material substance with the testing of the fitness and capacity of a human being is obvious, and the criteria of physical tests can and ought, with certain important limitations and extensions, to be applied to examinations of human beings. The principal criteria for physical tests may be stated as follows :

(a) The test used of a given characteristic should demonstrably yield a suitable measure of that characteristic. This is not a self-evident proposition (thus it is practically impossible to measure directly such a property as "absolute temperature," as defined in thermo-dynamics), and it is necessary to show how far indirect measurements are valid for their purpose (see under (ii), § 9, below). Many examination methods are indirect tests of the characteristics which it is desired to measure and compare.

(b) In cases where the whole of the material



## THEORY OF EXAMINATIONS

under investigation cannot be tested, it must be shown that the sample tested is an average sample of the whole.

(c) The test must be "trustworthy," *i.e.* it must, when repeated a number of times as nearly as possible under identical conditions, by the same observer, or when repeated by different observers, yield approximately the same results (the competence of the observers being assumed).

(d) The application of the test itself must not sensibly affect the characteristic tested, or injuriously affect the material as a whole.

9. For the sake of brevity we shall, in what follows, use the word "characteristic," as applied to examinations, in a wider sense than usual, to denote anything for which marks may be given to a candidate, *e.g.* knowledge of French, or of mathematics, capacity to ride a horse, etc. The numerical evaluation of human characteristics presents at least two kinds of difficulty which exist only to a much less extent with physical characteristics.

(i) Whereas we can in many (though not all) cases legitimately represent the values of physical characteristics either in terms of a unit, by straight lines of different lengths, or by numerical ratios, such representation in the case of human characteristics, though convenient, is often of doubtful validity. There is no such thing as "unit knowledge of mathematics," or "unit knowledge of French," nor could we justify a

## THEORY OF EXAMINATIONS

statement that "candidate A knows twice as much French or mathematics as candidate B." We shall deal further with this matter below.

(ii) The variation of human characteristics both with lapse of time and with the conditions under which human beings are normally placed is much greater, and less easily determinable than the variation of the physical characteristics of material substances in daily use concurrently with the variations of the physical conditions under which they are usually placed. Without going into an elaborate analysis, we may say that, while it is useful for practical purposes to say that the density of a piece of iron is 7·8, by which we mean that it differs little from that number under every-day conditions of temperature and pressure, it is more difficult to make a corresponding rough-and-ready statement in regard to the important mental characteristics, either general, such as "power of attention," or specific, such as "power to give an account of the Norman Conquest," since these characteristics vary under every-day conditions of interest and fatigue, and the law of such variations escapes definition.

10. It is clear that the knowledge of a characteristic of a purely evanescent kind is of little value. A candidate who knows the dates of the accession and death of the English sovereigns merely as a series of numbers on one day may have forgotten them on the next. Unless, there-

## THEORY OF EXAMINATIONS

fore, there is collateral evidence of the likelihood of the permanence of the knowledge itself, the only permanent or semi-permanent characteristic of which such knowledge shows evidence is the power to memorise numbers, and if we are ignorant as to the time taken to memorise them the characteristic is one to which we are not entitled to attach any great value on the evidence shown.

We may add to the criteria of examination tests enumerated above the following: (e) the characteristic tested should be demonstrably of a permanent or semi-permanent nature.

It frequently happens that, while examiners nominally assign marks for knowledge, they are perfectly aware that the knowledge may be evanescent, and they really intend to assign marks for what is assumed to be the permanent power of acquiring knowledge.

We shall use some or all of the above criteria in dealing with a few of the more important problems which arise in connection with examinations.

### V. THE PROBLEM OF PURPOSE

11. Before dealing with more special problems, we have to deal with the general problem of purpose, the most important, but, in very many cases, the most neglected of all examination problems. The majority of discussions about particular examinations, other than final professional examinations, are confused because the

## THEORY OF EXAMINATIONS

object and purpose of these examinations is not defined, and in the case of "pass" and "honours" examinations because the examiners do not tell us what a person who just "passes" or just gets "third" or "fourth-class honours," *knows or can do*. From experience we know that we cannot say with any certainty that a person who "passes" in a given subject will even be able to pass in that subject a year later. Probably if the examiners were pressed they would admit that "passing" the examination means nothing more than that a certain candidate was examined by certain examiners and was able to earn a certain percentage of marks in a certain number of subjects at a certain date. We have therefore next to investigate the problem of marks and marking.

### VI. THE PROBLEMS OF MARKING

**12.** Most examinations are made up of a series of tests, either of the same or different kinds. A number of marks is assigned to each question in each paper, or *viva-voce*, or practical examination set, and the maximum obtainable is the sum of the marks so assigned.<sup>1</sup> The total number of marks allotted to each candidate is the sum of the marks assigned to his several answers to the questions which he is asked to answer. The

<sup>1</sup> If a candidate is required to answer only a certain fraction of the whole number of questions, the maximum is reduced accordingly.

## THEORY OF EXAMINATIONS

ordinary working hypothesis adopted is that the value of the performance of the candidate is proportional to the number of marks he obtains, and that we shall be doing no injustice to the candidate as a whole, if we arrange them in an "order of merit" (better expressed as *order of proficiency*) in accordance with the numerical sums of such marks. In pass examinations a candidate is ordinarily required to obtain from 30 to 50 per cent. of the grand maximum assigned to all the tests which he is required to take, in order to be allowed to "pass."

To satisfy criterion (c) § 8 above, we should be able to show (i) that at a given examination a number of examiners, regarded as equally competent, will assign approximately the same relative mark-values to the different questions; (ii) that the same, or another set of equally competent, examiners will attribute the same marks to the several answers to those questions given by any candidate.<sup>1</sup> Further, if we are to be able to compare the proficiency of candidates of any year with the proficiency of those of a preceding year, we must have some adequate comparison of the "standard" at the two examinations in question, an exceedingly difficult and complex problem, of which nothing but a rough-and-ready solution

<sup>1</sup> And that the same examiner will allot the same marks if he marks the papers a second time. In examinations with hundreds or even thousands of scripts to mark, the question of the examiner's fatigue becomes of importance. (See Note E, p. 111, below.)

## THEORY OF EXAMINATIONS.

can be provided. It has been proposed in examinations at which large numbers of candidates present themselves to ensure a constant standard of "passing" by allowing the same percentage to pass each year, thus assuming the variations of marks obtained to be due to variations in the standard of the difficulty of the questions and of the marking of the examiners, or both.

13. The published evidence in regard to constancy of marking, notwithstanding its fundamental importance, is exceedingly slender. Edgeworth in 1888-90, in an investigation on First Class Civil Service Examinations, found the element of chance in the marking to be so great that only a fraction of from a third to two-thirds of the successful candidates could be regarded as above the danger of coming out unsuccessful if a different set of equally competent examiners had been appointed.<sup>1</sup> The element of chance may be expected *a priori* to vary with the purpose of the examination and to be least in a "final" professional examination, where, in awarding marks, the examiner has more or less consciously in his mind the relative importance of the tests in relation to the requirements of the profession and the fitness of each candidate to enter the profession. Where, as in a pass examination in "arts" or "science," the purpose of the examination is not defined, the meaning of a "pass"

<sup>1</sup> See Note E, p. 129, below.

## THEORY OF EXAMINATIONS

is necessarily indeterminate. Corresponding to this indetermination we find remarkable variations in the percentage of candidates who pass from year to year. It may be safely assumed that these variations are not due either to variation in the percentage of marks required for a pass, or to carelessness on the part of the examiners. Any variation in the percentage of marks required for a pass is probably made with a view to reducing rather than to increasing differences in the percentages of "passes" from year to year; and carelessness may easily tell in favour of as well as against any particular candidate.<sup>1</sup> Without careful statistical investigation on a large scale it is impossible to state what are the precise facts of the case or the precise causes of the variations of passes. They may be complex, and due (a) to real variations in the average proficiency of the candidates, or (b) to variations in the methods of setting papers and of marking the answers by the examiners, or (c) to a combination of these causes.

14. We have assumed, so far, that the total marks earned by a candidate are the sum of the marks earned on the different parts of the examination, but it is shown by experience that examiners are frequently not satisfied that the totals of marks assigned in the way described represent

<sup>1</sup> In well organised examinations special precautions are taken to eliminate error through marks being wrongly added or overlooked.

## THEORY OF EXAMINATIONS .

the proficiency of the candidates, and are apt to say "the candidate is a good mark-getter, or a bad mark-getter as the case may be."<sup>1</sup> Professor Bryan, F.R.S., as a remedy,<sup>2</sup> has proposed to abolish absolute maxima and to leave examiners free to add on marks for *exceptionally* good answers. The same result might be more conveniently achieved by assigning (say 20 per cent.) marks for general impression, so that the total marks might be made to correspond to the "resultant impression" of the examiner, which is probably, though unanalysed, a safer guide than the mere addition of marks, in the existing absence of statistical information as to the general trustworthiness of the marking system. Some examiners give negative marks for answers showing want of comprehension, but the practice is not general. Experience seems to show that examiners are more frequently in agreement as to the "class" of a candidate than as to his marks, the number of "classes" assigned being not more than four. In the minds of examiners there sometimes appears to be a real if undefinable distinction between the classes of two candidates, A and B, who have obtained the same numerical marks in the first instance; they may say A is a "first class" man, B is a "second class" man; and the separation of the two may be finally effected by an adjustment of marks.

<sup>1</sup> Stanley Leathes, *What is Education?* p. 154.

<sup>2</sup> *Loc. cit.* See note, p. 19.



## THEORY OF EXAMINATIONS

Possibly it might be shown that examiners in this way distinguish between evidence of "capacity" and of "memory" or knowledge.

15. In the "Greek-letter" system of marking each answer is marked  $\alpha$ ,  $\beta$ ,  $\gamma$ , or  $\delta$ , in the first instance, with (say) four further distinctions (+, + +, -, - -) attached to each of the first three letters. The system has some advantages, but it is difficult to see how it can be applied to a competitive examination on a large scale unless the symbols are translated into numbers.<sup>1</sup> It has been suggested that the only satisfactory way of determining "order of proficiency" (or "merit") is to compare each candidate's performance with that of each of the others and to assign a final order in accordance with the number of superiorities shown by the several candidates. The process is too laborious to be applicable to any large number.

16. Some of the greatest anomalies in marking arise from the failure to distinguish the principles applicable to (a) examinations designed to test the progress of a pupil, from those applicable

<sup>1</sup> Certain examiners assign Greek letters to individual answers in the first instance, and then convert them into numbers which are added up, so as to give a total for each individual. Thus, if the system indicated in the text is used, we have sixteen different symbols, to which are attached the successive numbers zero to 15. There is probably a real psychological difference between this process and what might seem to be the equivalent process of assigning 15 marks to each answer.

## THEORY OF EXAMINATIONS

to (b) examinations designed to test practical capacity. Take the case of simple addition, and suppose that the candidate can only do two addition sums right out of six. If the examiner is only testing the progress of the candidate, it may be legitimate to allot to him 30 per cent. of the marks; but, if the examination is designed to test his power of adding correctly, he should receive no marks, since that power has not reached the point of practical usefulness. The same kind of argument is applicable to all kinds of capacity tests; for these there should be no mark allotted in examinations in category (b) above unless the candidate obtains considerably more than half (say 75 to 90 per cent.) of the marks assigned to the particular test; but the argument does not apply to memory tests, in which a candidate is legitimately entitled to marks for each item remembered. •

17. In what precedes we have preferred "order of proficiency" to "order of merit" because in the anonymous examination, as ordinarily conducted, we know nothing of the "merit," or deservingness, from the moral point of view, of the candidate.<sup>1</sup> Speaking roughly, proficiency, as shown in an examination room, may be said to depend *inter alia* on three variables, (a) natural ability, (b) opportunities (including opportunities of getting good teaching), and (c) industry and perseverance. With a particular purpose in view,

<sup>1</sup> See Note G, page 138.

## THEORY OF EXAMINATIONS

it may be justifiable to place A, who is a clever, lazy dog, above B, who is less clever but more industrious ; with another purpose in view, the order ought to be reversed. In considering a permanent appointment to any post, it is clear that examination results taken by themselves are insufficient.

When the results of a competitive examination can be made to depend in part on the report of trustworthy persons familiar with the previous history of the candidates some account may be taken of the relative importance of the three factors (*a*), (*b*), and (*c*), in contributing to the proficiency shown, and weight may be given to each factor corresponding to the general purpose of the competition.

### VII. PROBLEM OF THE EXAMINATION-SAMPLE

See criteria (*b*), (*c*), and (*e*), §§ 8 and 10 above

18. Examinations may be roughly divided into two classes, according as (i) the examination itself covers the whole field to be tested and the results give complete information as to the characteristics required,<sup>1</sup> or (ii) the field of characteristics can only be tested by "sample." We can show with sufficient completeness by examination that a candidate has the power accurately to add, divide, or multiply, to read a thermometer, or

<sup>1</sup> To simplify the problem, we neglect here the variability of condition of the candidate (see § 22 below).

## THEORY OF EXAMINATIONS

measure an electrical resistance, the power to write in his mother-tongue clearly, grammatically, and in accordance with a logical plan, to speak a foreign language, to write shorthand, and a great many other things of primary practical importance.

If the examination be sufficiently complex we can, as in the case of actuaries, cover the whole field of a professional examination, so as to be able to say with certainty, this candidate can do everything required of an actuary. But the majority of examinations test only by sample. As a typical example of cases where judgment by sample fails, we may quote that of an examination in "unseen Latin" in which the examiner happens to pitch on passages with which the candidate is acquainted. The conclusion drawn in such a case is obviously wrong. Even such complex professional examinations as those in medicine cover only a small portion of the field.

19. The question of "fairness of sample" only approaches to definiteness when the field is clearly defined either by the well-known requirements of a profession or by examination syllabus. Otherwise the field itself will vary necessarily with the examiner, and the comparative success

<sup>1</sup> The fact of failure to pass such an examination does not necessarily prove that a candidate would not show the characteristic tested, if he were placed under conditions other than those of the actual examination (see § 22 below).

## THEORY OF EXAMINATIONS

of two candidates equally well equipped in every way will depend on whether their views (or those of their teachers) of what is important coincide or not with those of the examiners.

**20.** Let us consider, first, professional examinations by sample. The professional examinations which the public probably regard as of the greatest importance are those in medicine. Here the field of minimum requirements, though constantly varying, is fairly well defined at any given time. In order to make sure that a candidate shall not succeed by *chance*, the public insist that he shall have gone through a definite course of instruction certified by his teachers, in addition to passing the examination conducted either by his teachers, or by "external examiners," or by a combination of the two. In most other cases public opinion is content with less stringent guarantees to provide against the possibility of success by chance, and to ensure that the examination-sample shall be a fair sample of capacity and knowledge.

**21.** In the case of non-professional examinations by sample, *e.g.* honours examinations in history or literature, the element of chance is greatly increased, especially when the examiners have no knowledge of the previous training of the candidate; and where such examinations are competitive the element of chance would appear seriously to affect the validity of the result (*cf.* § 13 above).

## THEORY OF EXAMINATIONS

**22.** The question how far a particular candidate is in a normal condition during the examination is a special aspect of the problem of sample.

(1) The candidate may, by chance, be in an abnormal condition, due to temporary indisposition. It is impossible in most examinations to make any correction to allow for this; yet failure in the case of a candidate who is ill at the time of the examination gives no evidence as to his powers when he is well.

(2) We know that capacity-characteristics are far more permanent than memory-characteristics; they imply habits of thought, not easily destroyed or lapsing with time. From one point of view memory-characteristics may be regarded as evidence of the capacity to memorise, but the ordinary examination gives no information as to the time which has been required to memorise in a particular case. It is only in the case of examinations of extreme difficulty, like the Class I. Civil Service examinations, with an age limit, and covering a very wide field, that we get valuable evidence of the power to memorise rapidly.

(3) The candidate may be in an abnormal condition due to the examination itself. Some candidates are stimulated by the excitement of examinations to a display of powers of which they are not normally capable; with the majority, owing to nervousness and fatigue, the reverse is probably true.

The question of fatigue is one which demands

## THEORY OF EXAMINATIONS

investigation. Does, for instance, a written examination conducted for six hours a day on five successive days diminish the average mental powers sensibly, and, if so, does it diminish them in approximately the same proportion for all candidates? Do we definitely wish to eliminate certain candidates by this fatigue-test? Will the results in subjects requiring special concentration, *e.g.* mathematics, be affected according as the papers in those subjects are placed at the beginning or the end of such an examination?

(4) The question of "nervousness" specially affects *viva-voce* and practical examinations. The question whether allowance should be made for such nervousness at a *viva-voce* depends on the purpose of the examination. If it is definitely intended to test the candidate's power of "thinking on his feet" before an audience, no allowance should be made. If it is intended only as a test of knowledge or capacity in a special subject which will *not* be tested in the same way in the career for which the candidate is intended, the test itself may be a failure, if the candidate is nervous.<sup>1</sup> In many foreign examinations the *viva-voce* is definitely used as a test of power of exposition. The case of the practical examination is somewhat different, as practical tests have

<sup>1</sup> When the purpose of an examination is properly defined, it is easy to show in certain cases (as when it leads to the selection or passing of an unsuitable candidate) that the examination has failed.

## THEORY OF EXAMINATIONS

rarely to be carried out before a critical audience, and it is expedient if possible so to conduct practical examinations that a candidate shall not be able to fail entirely through a single act due to nervousness.

**23. Testing of Schools by Sample**—A special aspect of the problem of sample is presented when the efficiency of a school is judged by testing a certain proportion of its pupils selected either at random, or better, from a classified list drawn up by the school. The testing, by an external examiner, of a few pupils classified by the teachers as the best, the average, and the worst may be made to yield a fair comparison of his own judgment with that of the teachers.

### VIII. INFLUENCE OF EXAMINATIONS ON EDUCATION AND TEACHING

**24.** In 1889 the Hon. Auberon Herbert published a collection of opinions of distinguished people entitled, *The Sacrifice of Education to Examinations*, and Mr. Stanley Leathes has probably expressed what is still the popular view in regarding examinations as a "necessary evil" (*loc. cit.*, p. 141). But, as a matter of fact, the influence of the examination on the candidate, whether for good or bad, depends entirely on whether it is well or badly designed for its purpose, and the general question "Are examinations 'good' or 'bad'?" is indeterminate,



## THEORY OF EXAMINATIONS

*i.e.* it cannot be answered in the negative or the affirmative. If it is essential for a candidate's future that he shall be able to speak French to strangers at a given age, his education up to that age will not be sensibly influenced by the fact that he is required to pass a *viva-voce* examination in French ; similarly with the examination for actuaries (see *supra*). Here the influence of the examination, considered by itself, is nil. The nature of the examination and the previous education are alike determined by a consensus of opinion, and the examination *qua* examination does not influence the candidate or his teachers for good or for evil. The fact remains that in many cases there can be no such consensus of opinion ; and in these cases the examination may exert an independent effect, prejudicial or beneficial, or a combination of both, on previous education.

It is in the nature of things, especially in non-professional examinations, that there should be great diversity of opinion both as to the proper field, and even the purpose, of a particular examination, and that diversity will become most marked in those branches where progress is being made and recent developments appear in rapidly changing perspective. The purpose of an examination being defined, it is open to question, in any given case where there is disagreement whether the view of a particular examining body or whether the view of a particular teacher is

## THEORY OF EXAMINATIONS

best adapted to the fulfilment of that purpose.<sup>1</sup> But it is certain that where there is such conflict of view either the candidate's teaching must be subordinated to the views of the examining body, or he must risk failure merely because he is being trained with one purpose in view and tested with another. If the teacher examines or has a sufficient share in the design of the examination the alternative of conflict or subordination does not arise. Where it does arise the personal efficiency of the teacher is bound to be diminished if he is forced to teach in a way which he regards as second best. Professor T. P. Nunn has pointed out, on the other hand, that examinations serve in certain cases to stimulate teachers, but those cases are limited to certain categories. It is clear that where teachers have some share in the examination of their students the examination syllabus is more flexible and can be more readily changed from year to year. (For question of anonymity and fairness of examinations conducted by teachers see § 29, below).

**25. *Effect of Examinations on Powers of Exposition.***—In ordinary life a sensible person only speaks or writes on any subject for the benefit of a person who knows less than himself; the

<sup>1</sup> The independent judgment of the candidate himself, important in questions of culture, is only in the rarest cases likely to be equal or superior to that of examiners or teachers in other matters, and the sacrifice of the candidate's own judgment in training for examinations is generally inevitable (see *infra*, § 27).

## THEORY OF EXAMINATIONS

examination candidate writes for a person who, in general, already knows what he has to say, *i.e.* the examiner, and hence may be trained unwittingly to express himself obscurely and by allusion. It is of great practical importance to counteract this powerful tendency by rational training in composition.<sup>1</sup>

Sir Napier Shaw thinks that the actual form of the three-hours' papers now set has exercised an adverse influence on the teaching given in physical science.<sup>2</sup>

**26. Effect on Powers of Concentration.**—In the case of many candidates, examinations, through the fear of failure, induce concentration during working-time which it is difficult to secure in any other way.

<sup>1</sup> Cf. *The Writing of English*, by the present writer, pp. 73, 74 and *passim*.

<sup>2</sup> "An evil microbe has got hold of the teaching of physical science, for I know of my own knowledge that some at least of the examiners outside the examination-room take a real interest in things. I attribute the sawdust with which the examinations are stuffed to the influence of [a] prolonged succession of examinations of the same type, the three hours' written form of exercise which corresponds with no problem that any man or woman has to face in practical life, except the single one of keeping awake while the examination lasts. Why can one not invent some other form of examination, one, for example, in which candidates are asked, as they are in practical life, to take up a job and say when they have finished it?"—*The Lack of Science in Modern Education*, by Sir Napier Shaw, Sc.D., F.R.S., Director of the Meteorological Office (Lamley & Co., 1916), pp. 13, 14. The pamphlet is in part a plea for the study of "real things," as opposed to "laboratory science."

## THEORY OF EXAMINATIONS .

**27. *The Effect on Education of Examinations designed to test Culture.***—If culture be distinguished clearly from erudition, and be interpreted as an individual sensitiveness and responsiveness in regard to those things which can be neither weighed nor measured but are nevertheless regarded as of value by the community,<sup>1</sup> then it may well be held that culture cannot be tested by any formal examination, but only by social intercourse, or by the evidence of original work (thesis or dissertation). A written examination can test a knowledge of literature, or a knowledge of history, it cannot test the literary or the historical sense. But the training for such examinations, by checking the individual impulse of the candidate at every turn, may well kill the individual culture and originality of effort which it cannot test. The sensitive inquiries of the examination candidate are likely to be checked at every turn by the inevitable question—Will they “pay” at the examination?

**28. *The Effect of Examinations on Education by Over-absorption of Time.***—If examinations cannot test culture, still less can they test (except in certain limited ways) character, or public spirit. It is therefore of the first importance that the curricula of candidates, and especially of candidates in secondary schools, shall be designed so that training in the subjects tested by examinations shall only absorb a portion of the candidate's

<sup>1</sup> See Notes C and D, pp. 93 and 95 below.

## THEORY OF EXAMINATIONS

time. In exceptional cases the effect of preparation for examinations is to ruin a candidate's health by over-work. The effect of examinations on the health of girls and the precautions necessary to avoid overstrain in their case deserve special consideration.

### IX. IMPARTIALITY AND FAIRNESS OF EXAMINATIONS

• 29. As applied to examinations, the terms impartiality and fairness are not synonymous. In cases where the results of examinations are competitive, it is of the highest importance that they should be, as far as possible, not only impartial, but *fair* in the sense that the candidates shall be really arranged in order of suitability for the purpose for which the examination is held.

Examinations in which the candidates are entirely unknown to the examiners are necessarily impartial, but the impartiality may be to a considerable extent that of chance, and have all the injustice of chance. To have anticipated, or to have been coached, in a number of memory-tests which are later selected by the examiners may bring success, but it affords no proof of superior suitability. It is to be pointed out that anonymity can only be fully preserved when an examination is wholly written ; as soon as the candidate is brought face to face with the examiner at a practical or *viva-voce* examination, the anonymity

## THEORY OF EXAMINATIONS

may disappear. The more searching and personal the test, the greater are the possibilities both of fairness on the one hand and of partiality on the other, everything depending on the character and capacity of the examiners.

Possibly the most satisfactory method of selection is to make a first choice anonymously, by a written examination, and then to test the surviving candidates by *viva-voce* examination before a Board, either with or without evidence from those previously acquainted with the candidates. Such a Board may of course be prejudiced against particular candidates and categories of candidates, and against such prejudice there is no safeguard but public opinion. As an example of successful selection by boards of examiners may be quoted the method used by Lord Cromer in choosing officers for the Soudan.<sup>1</sup> The efficiency of such methods of selection is tested by the results. ‘

### X. CLASS I. CIVIL SERVICE EXAMINATIONS<sup>2</sup>

30. These competitive examinations are of a peculiar type. The candidates may select, with certain limitations, as many subjects as they like out of a large number (to which are allotted different maxima), as long as the total number of marks obtainable does not exceed a certain

<sup>1</sup> See p. 45 above.

<sup>2</sup> See also Introduction, p. xii, above, and Note A, p. 80, below.

## THEORY OF EXAMINATIONS

limit, and there is no necessary relation between the subjects which a candidate takes and his future career. It is clear that this is a case in which the purpose of the examination is not to test directly special knowledge of any kind, but to test the power to master a number of subjects ; the severity of the competition makes the preparation for the examination to some extent a test of the candidate's character, inasmuch as to ensure success he must for a time sacrifice all other interests to it. Success at the examination is followed by different requirements according to the service entered by the candidate. Those selected for the Home Civil Service go through a period of probation without further examination ; those selected for the Indian Civil and certain Colonial services have to undergo further probationary examinations in special subjects required for their official work.

A fundamental criticism of the methods adopted at these examinations is that it is impossible to compare the suitability of the candidates by means of heterogeneous aggregates of marks assigned by the different examiners in different subjects. Since on the whole the system has worked well it may be assumed that the number of suitable candidates is considerably in excess of the number selected in each year. It is commonly stated that marks may be " more easily earned " in subject A than in subject B, but the validity of such a statement in respect of different

# THEORY OF EXAMINATIONS

pairs of subjects will almost necessarily vary from year to year.

## XI. SUGGESTED USE OF EXAMINATIONS TO TEST COMMON SENSE AND POWER OF JUDGMENT

31. Hitherto comparatively little use has been made of examinations to test common sense and power of judgment in a way for which skilful preparation could not furnish any illusory substitute. The writer has suggested that "English" papers may be set of a kind to test "common sense, intellectual tact, readiness and order."<sup>1</sup>

## XII. EXPERIMENTAL PSYCHOLOGICAL TESTS

32. A large amount of important work has been done recently by psychologists with a view to determining the natural capacities of individual "educands" by experimental methods. Prof. C. Spearman and Dr. F. H. Hayward have suggested that experimental methods may replace or coalesce with the examination methods hitherto used. In America experimental methods have been successfully applied by F. W. Taylor, Münsterberg, and others to determine relative fitness for various kinds of industrial work. The methods seem capable of great development;

<sup>1</sup> *The Writing of English*, pp. 81-4, and paper on Examinations, read before the British Association, published in the *School World*, October 1911; the new English papers proposed by Mr. Stanley Leathes' Committee for Class I. Civil Service Examinations should serve this purpose (see p. 92).



## THEORY OF EXAMINATIONS

and a system of tests is being used in American "vocational bureaus" for the purpose of advising youths in the selection of a career. It is possible that the whole system of examinations may be ultimately transformed by the use of these quantitative methods; but the matter is still in a rudimentary stage.

### XIII. MEDICAL TESTS

• **33.** For certain careers a medical test of fitness is essential. There is a movement on foot to extend medical tests and to give marks in University degree examinations for physical capacity as shown in Swedish drill and the gymnasium, a return to Greek ideals.

### XIV. CONDUCT OF EXAMINATIONS

**34.** To avoid chance hardships many precautions of detail are necessary in the conduct of examinations, of which the following are a few examples, taken at random as illustrations.

(1) It is the practice of some examiners to attribute different mark-values to different questions, leaving it as an additional test of the candidate's intelligence to determine in what way they can earn the largest number of marks within the period of the examination. Definite instructions on this point and on the minimum number of answers which must be attempted to ensure a

## THEORY OF EXAMINATIONS

pass should be given if the examining body holds a particular view on the matter.

(2) In the case of mathematical papers the examiners should work out the questions in full, to avoid unexpected difficulties and even "traps."

(3) It is essential that examiners should not be required to mark an excessive number of scripts in a given time. (See p. 110.)

### XV. CONCLUSION

35. In what precedes a method has been sketched out and illustrated, for dealing with examination problems. But the question is only in its infancy. A systematic investigation (including statistical investigation) of the method and results of examinations is urgently needed. A great advance would be effected if examining bodies could be induced to express in words (1) the purposes of their various examinations, and (2) as exactly as possible what the "passing" of a particular examination means in terms of what all the successful candidates could certainly do at the time of the examination—in other words, non-professional examinations should be given the same precise significance as professional examinations.

---

## NOTE A

### EXTRACTS FROM THE FOURTH REPORT OF THE ROYAL COMMISSION ON THE CIVIL SERVICE (MAJORITY REPORT), 1914.<sup>1</sup>

#### (1) *First two paragraphs of Section 7 (p. 29)*

7. But before proceeding to discuss in any detail the schemes of examination by which we propose that those tests should be carried out, we desire to emphasise our opinion that an examination scheme is only part, and not the most important part, of the machinery by which the State should aim at the selection of suitable candidates for its service.

Unless the general educational system of the country is efficient, the natural powers of those who would make the best candidates will neither be discovered nor trained.

#### (2) *Third paragraph of Section 8, and first two paragraphs of Section 9 (p. 31)*

8. In Chapter I we have indicated the origin and followed the development of the competitive system of recruitment for the Civil Service. With the progress of this development the efficiency of the Service

<sup>1</sup> [Cd. 7338.] Published by the Stationery Office. Price 1s. 4d., to be purchased of any bookseller.

## EXAMINATIONS—NOTE A

has grown until at the present time the State possesses a body of public officers who are far more competent and zealous than their predecessors, appointed under the regime of patronage, are stated on official authority to have been. We have no doubt whatever that to this highly satisfactory result the system of competitive examination has mainly contributed. The system has, in our opinion, entirely justified the expectations of its originators. It is true that it never has been, and is not now, exempt from hostile criticism, and we do not affirm that the written examination is an infallible or a final test of the best results of education. But the defects attaching to the system, which we shall in due place examine, are, perhaps, susceptible of some mitigation. In existing political conditions, and in such developments of them as can be reasonably anticipated, we believe that the advantages of the system of competitive examination as a means of recruitment for the Civil Service far outweigh any defects which have come to our notice, and we are convinced of the importance not only of adhering to the system, but of extending it whenever possible.

9. Since, then, examination should, in our opinion, play so large a part in the selection of this most important body of officials, it becomes of supreme importance so to order and arrange the schemes and improve the methods of examination as to secure the best results.

How far the existing arrangements secure this will appear from a review of the schemes of examination by which at present access is gained to the different administrative and clerical classes. We propose

## EXAMINATIONS—NOTE A .

now to undertake such a review, and to recommend such changes as appear desirable.

### (3) *Sections 26 and 27 (pp. 34, 35)*

26. We have already drawn attention to what is, in our judgment, a serious defect in the existing arrangements of the examinations for entry into the clerical branch of the Civil Service, viz. their divergence from the actual conditions of education. The Second Division examination marks the extreme point of such divergence.

The three years seventeen to twenty correspond to no stage in the organisation of education in this country. As a complete course of secondary education at present ends when the pupil is about eighteen years of age, it is obvious that, while the inferior age limit is now too low, the superior limit is far too high.

27. An inevitable, and in our judgment undesirable, consequence is the prevalence of "special preparation" or "cramming" for this examination.

We have taken pains to ascertain the extent to which cramming prevails, and we find that more than half the candidates undergo some kind of special preparation for this examination. To some extent recourse to the "crammer" is explained by the circumstances of boys already in employment during the day who have only the evening for study, and for whose needs no other agencies suffice. We attach greater significance, therefore, to the statistics showing, for candidates who give their whole time to instruction, the percentage who attend places of "special preparation."

Such statistics show that the number of candidates

## EXAMINATIONS—NOTE A

who leave school for this purpose is greater between 15 and 16 than between 14 and 15 years of age; greater still between 16 and 17, and greatest between 17 and 20.

This tendency is doubly objectionable. In the first place such special preparation necessarily costs money, while those boys from the elementary schools who are likely to be successful in this examination would, in most cases, be receiving gratuitous instruction in the public secondary schools. Moreover, in so far as this system makes success dependent upon an expensive substitute for the ordinary school education, it necessarily places the poorer candidates at a disadvantage.

In the second place, the interruption of a school curriculum is injurious to the boy's best interests.

More general recognition is now being won by the great principle that the most valuable elements in education are those formative influences which help to mould and develop character. Such influences form no part of the atmosphere of a "cramming" institution, the main object of which is to surmount or evade the difficulties of an examination paper with the smallest expenditure of time and trouble.

We do not deny that useful instruction may be given in such institutions. In so far as they supplement the strictly educational curricula of the schools of the country by teaching such technical subjects as book-keeping, shorthand, and typewriting, they perform a useful service to the public; but their professed aim is different from the aim of true education. The principal aim to be kept in view in recruiting the Civil Service (towards which examina-

## EXAMINATIONS—NOTE A

tions are a necessary, if a somewhat fallible means) is to obtain the best results of the regular educational system which is in actual operation ; and, when that system becomes sufficiently uniform throughout the United Kingdom to permit such a course, we think that Government should require a school-leaving certificate as a condition of candidature for the clerical examinations.

(4) *First two lines of the fourteenth paragraph of Section 43 (p. 40)*

43. It is to be remembered that competitive examination is a means, though necessarily an imperfect one, of testing natural ability, as well as acquired training.

(5) *Section 44 (pp. 40, 41)*

44. In regard to the suitability of the present Class I examination, academic opinion was not unanimous. There was, indeed, something approaching agreement in the view that the competitive examination is the most, if not the only, practicable means of selecting candidates for the situations which we are now considering. But the imperfections of a written examination as a test, not merely of knowledge, but also of capacity—of the faculty of sound judgment, common sense, resourcefulness, and resolution—were often insisted upon.

To make good these defects several suggestions have been put before us, to two of which in particular we invite attention, viz. : (a) the selection by the authorities of the various Universities of Honours men, also distinguished for their judgment, good sense,

## EXAMINATIONS—NOTE A

and general capacity in student life, for competition *inter se* for the situations vacant ; and (b) the institution of a *viva-voce* examination supplementary to the written test, with the object of ascertaining the candidate's general mental calibre, and the possession of those qualities of common sense, sound judgment, resourcefulness and resolution, upon which the written paper rarely gives assurance.

The Vice-Chancellors of the Universities of London and Wales entertained a favourable opinion of the former of these suggestions ; indeed they seemed to regard it as likely to become the normal method of recruitment of the future. We have considered whether it could be applied as an experimental measure to a proportion of vacancies annually, but we are satisfied that the practical difficulties are too serious to permit even a partial or tentative trial of the system. We are unanimously of opinion that, whatever the future may have in store for us in this connection, the method cannot generally be applied to-day, if only in view of the impossibility of closing the competition to non-university candidates.

(6) *First paragraph, and first eight lines of the second paragraph of Section 45 (p. 41)*

45. In respect of the second suggestion referred to above, we call attention to the opinion which the Civil Service Commission entertained on this matter in the earlier years of their activity. In their Tenth Report (1865) they say of the examination for the Indian Civil Service :—" In each subject there is a *viva-voce* examination, to which we attach great importance, introduced for the express purpose of fully



## EXAMINATIONS—NOTE A

testing the genuine character of a candidate's knowledge, in addition to calling into play those qualities which are little, if at all, elicited in an examination by written papers, viz. his readiness, self-reliance, and moral courage." We have ascertained from the Civil Service Commission that this practice of *viva-voce* examination was continued until the year 1891, when it was abandoned, apparently owing to the increase in the number of candidates caused by the raising of the age limits.

We do not now propose for the Home Service the establishment of the *viva-voce* examination conducted by each examiner in each subject taken by a candidate, although some of us believe that such an examination would be an effective help towards determining the candidate's knowledge of the subject. Our object is different : it is to ascertain whether a general *viva-voce* examination based on the whole syllabus, and conducted by a single committee of three carefully chosen examiners, might not be made an effective help towards determining those of the candidate's qualities which are little, if at all, tested in an examination by written papers.

### (7) Section 46 (p. 41)

46. Certain academic witnesses seemed to think that the present syllabus gives undue weight to the studies most commonly followed in the older Universities and that the modern institutions are handicapped by it in the competition. Undoubtedly the most salient fact emerging from an analysis of the academic antecedents of the successful candidates is the great numerical preponderance of those who

## EXAMINATIONS—NOTE A

have been educated at Oxford and Cambridge, especially at Oxford.

(8) *Section 50* (pp. 42, 43)

50. However conclusive in the opinion of some, and probable in the opinion of others, may be these explanations of the facts mentioned in paragraph 46 above, they do not in our judgment offer a sufficient reply to the inquiry whether the conditions of the examination, the grouping of the subjects, and the relative values assigned to each group, do or do not give to any subject or course of study an undue advantage in the examination hall. . . . It is, in our opinion, highly undesirable that such an inquiry should remain unanswered. A satisfactory answer to it can be given in only one way, namely, by a full investigation into the facts. We were glad to learn from the Civil Service Commissioners themselves that they concur in the necessity for such an inquiry ; and we accordingly recommend the Government to appoint a Committee with the object of ascertaining whether there is any foundation for or reality in the misgiving which certainly prevails in some quarters that the scheme of examination for the Class I. unduly favours the curricula of the older Universities and handicaps those of the newer. On the Committee the Civil Service Commission and both University and School experience should be represented, and it should be presided over by some man of recognised judgment and eminence in public life. Should it be found that any change is desirable, then it would be necessary for the Committee, while maintaining the high standard necessary for the examination, to revise and

## EXAMINATIONS—NOTE A

rearrange the syllabus, weighing, with even balance, the educational values of the studies of the older Universities against those of the modern learning, as it has been called, to which the younger Universities are thought to pay more attention. The revised syllabus thus made out should be submitted to the Government with whom the decision must rest.

(9) *Sections 4 and 5 of the Summary of Recommendations* (p. 101)

The General Civil Service, Administrative and Clerical.

(a) *General*

4. The competitive examinations for recruiting each class of officer, administrative and clerical, should be adjusted in respect of the age of competitors and the subjects of competition to the stages of the educational system actually existing in the country. (Chapter **II**, § 6, p. 29.)

5. The examinations should be directed to testing the natural ability of candidates, and the results of their education both with respect to acquirement of knowledge and the formation of mind and character.<sup>1</sup> It should not be directed to testing proficiency in particular subjects which lie outside the normal scope of education.

<sup>1</sup> The Commission give no hint to the Civil Service Commission as to how examinations are to be directed to testing the formation of character.—P. J. H.

## NOTE B

### ON THE REPORT OF THE TREASURY COMMITTEE ON CIVIL SERVICE, CLASS I. EXAMINATION (1917) <sup>1</sup>

THE Committee referred to in the heading was constituted as follows: Mr. Stanley Leathes, C.B. (Chairman); Sir Alfred Ewing, K.C.B., F.R.S.; Sir Henry Miers, D.Sc., F.R.S.; Mr. H. A. L. Fisher, F.B.A. (replaced on his appointment as President of the Board of Education by Dr. W. H. Hadow, M.A.); and Prof. W. G. S. Adams, M.A., with Mr. D. B. Mair, M.A., Director of Examinations to the Civil Service Commissioners, as Secretary. The reference to the Committee was:

To consider and report upon the existing scheme of examination for Class I. of the Home Civil Service:

To submit for the consideration of the Lords Commissioners of His Majesty's Treasury a revised scheme such as they may judge to be best adapted for the selection of the type of officer required for that class of the Civil Service, and at the same time most advantageous to the higher education of this country:

And in framing such a scheme, to take into account, so far as possible, the various other purposes which the scheme in question has hitherto served, and to

<sup>1</sup> [Cd. 8657.] Price 3d. Published by H.M. Stationery Office, and to be purchased of any bookseller.

## EXAMINATIONS—NOTE B

consult the India Office, the Foreign Office, and the Colonial Office as to their requirements, in so far as they differ from those of the Home Civil Service.

The scheme of the Committee may be summarised briefly as follows :

The examination is to be divided into three sections: Section A, an examination mainly general in character, obligatory for all candidates, with a maximum of 800 marks; Section B, a specialised examination, including fifty-nine separate subjects with maxima varying from 100 to 400, candidates being allowed to take up subjects in the section up to a total of 1,000 marks; *Extra Numerum* Section, covering 100 additional marks for translation papers in a language not already taken by the candidate.

The main points in which the scheme differs from the existing scheme (September 1917) is in the character of Section A, and the readjustment of the scale of marks in Section B, which is intended to place on an equality "the chief studies which are pursued by students at the University up to an Honours Course, viz. Classical languages, history and literature; Modern languages, with history and literature; History; Mathematics; and the Natural Sciences."<sup>1</sup>

Section A includes the following items: (1) Essay (100 marks); (2) English (100 marks); (3) Questions on contemporary subjects, social, economic, and political (100 marks); (4) Questions on general principles, methods, and applications of Science (100 marks); (5) Translation from one of a number of

<sup>1</sup> To obtain the maximum in Natural Sciences, candidates must take one main subject up to the "higher level" and two subsidiary subjects up to the lower level.

## EXAMINATIONS—NOTE B

European languages, or Latin (100 marks); (6) A *viva-voce* examination (300 marks) Total, 800 marks.

"This Section," says the Report, "is intended to test the candidates' knowledge of the English language and their capacity for its skilful use, their accurate command of knowledge which they should have acquired in the course of a systematic education and self-education and should have retained to assist them in their future work, and their equipment in one foreign language at least for working purposes. The languages selected are those most likely to afford information useful to public servants. As circumstances change others should be added at the discretion of the Civil Service Commissioners.

"The *viva voce* should be a test, by means of questions and conversation on matters of general interest, of the candidate's alertness, intelligence, and intellectual outlook, his personal qualities of mind, and mental equipment.<sup>1</sup>

"It is not intended that any candidate should be disqualified for failure in any of the parts of this Section, or in the Section as a whole, but that the Section should count substantially in the competition."

The Section is devised to ensure, as far as possible, that *general* knowledge of both science and the humanities which Huxley, a pioneer in this matter, pointed out as essential for every "well-educated man."

Before Huxley's time, education meant the humanities only; since his time some would have it mean "science" only, in a narrowly restricted sense. If

<sup>1</sup> [Cf., on this point, section 45 of the Majority Report of the Royal Commission on the Civil Service, 1914, quoted on p. 85 above.]

## EXAMINATIONS—NOTE B

the Committee's scheme is adopted, the Class I. examinations will at any rate be free from the reproach that they tend to perpetuate the "neglect of science" which has been so serious a reproach to our education, and which has had such grave consequences in the War. The Committee have given specimen examination papers; but the principles to be adopted in marking are only vaguely indicated by the following rubric: "Each candidate's answers will be valued as a whole, and any candidate who presents answers to less than three or more than four of the questions will be presumed to do so in knowledge of the value of the answers which he is able to give to the questions he chooses."

The specimen English paper, which is on lines largely in accordance with suggestions put forward at various times by the present writer, constitutes a new departure in examinations of this character. It is a test of general intelligence of which the result cannot possibly be falsified by special preparation.

The re-introduction of the *viva-voce* test, to which 300 marks are attached, would again, as I have suggested (pp. 21-22), constitute a real reform. But it is one which needs very careful safeguarding; and I am not sure that it would not be well to make it public, like so many foreign *viva-voce* examinations. The presence of strangers who could not influence the verdict would add little or nothing to the severity of the test, for which clearly the candidates would prepare themselves, by undergoing similar tests in class before their comrades.<sup>1</sup>

<sup>1</sup> Cf. "The Aim of the New Curricula of French Secondary Schools for Boys," by the present writer, in *Special Reports of the Board of Education*, vol. 24.

## NOTE C

### THE CONNOTATION OF THE WORD "CULTURE" (see also Note D)

THERE are many interesting passages bearing on this point in J. H. Newman's *The Idea of a University*, third edition (1873), especially in the "Discourse on Knowledge viewed in relation to Learning."<sup>1</sup> The following (p. 135) is typical of Newman's views :

"When this analytical, distributive, harmonizing process is away, the mind experiences no enlargement and is not reckoned as enlightened or comprehensive, whatever it may add to its knowledge."

Newman seems to have in his mind only that colder aspect of additions to knowledge by assimilation of fresh facts to pre-existing intellectual possessions which the modern psychologists call "apperception."

The following passage from Goethe gives a fuller description of the same process, in which the emotional aspects are taken into account :

<sup>1</sup> The book was originally published in 1852 under the title, *Discourses on the Scope and Nature of University Education*, etc., and the particular essay referred to was entitled, "On Philosophical Knowledge viewed in relation to Mental Acquirements."



## EXAMINATIONS—NOTE G

" Everything of outstanding excellence (*Alles Vortreffliche*) gives us for the moment a sense of limitation, because we feel we have not reached its level ; only in so far as we afterwards make it part of our own culture (*Cultur*) and an element in our own mind and spirit does it become to us dear and valued." <sup>1</sup>

There is an interesting parallel passage in E. F. Benson's clever contemporary novel, *The Climber*:

" . . . but that assimilation which is necessary before facts can become the food of culture, that kindling of the blood, as with romance, had not occurred."

I do not quote from Matthew Arnold's *Culture and Anarchy*, partly because it is so well known, partly because I propose to deal with some aspects of it elsewhere.

In the text I have compared the individuality of culture with the individuality of conscience, a comparison suggested to me, I think, by an interesting book by the French physicist, Henri Bouasse (whose constructive suggestions seem to me more valuable than his criticisms of his fellow-workers), but the comparison ought not to be pushed too far. Culture is clearly to be distinguished not only from the moral but from the " intellectual " conscience to which I have referred in the *Writing of English*, p. 63.

<sup>1</sup> *Maximen und Reflexionen*, Goethe Gesellschaft edition, 1907, p. 65.

## NOTE D

### KULTUR AND CULTURE

THE history of the word *Kultur*<sup>1</sup> in Germany is from the point of view of pure semantics almost a paradox; from the point of view of the history of mankind it is grave and tragical—*corruptio optimi pessima*.

Thanks to successive editions of the Dictionary of J. C. Adelung, we can safely say that the word became of general use in Germany between 1774 and 1793. It was introduced from the French, and took with it the two active significations, the literal one of cultivating the land, and the figurative one of cultivating the mind, without any marked departure from the French original.<sup>2</sup> But in the figurative sense in which the word is most generally employed, to signify the *result* of such cultivation as applied to man, there is a wide and important departure. Whereas "culture," both in French and in English, is always used with special emphasis on the intellectual and

<sup>1</sup> *Cultur* was originally spelt with a C, and is so spelt in most of the earlier texts. J. H. Campe, however, in the posthumous edition of Adelung's Dictionary points out that the letter C is unnecessary in German, as it may always be replaced by K (as in *Kultur*), or by Z; and this method of spelling is now in general use in Germany.

<sup>2</sup> Cf. the definition of *culture* in the edition of the *Dictionary of the French Academy* of 1762.

## EXAMINATIONS—NOTE, D

æsthetic aspect of its general connotation (Murray's Dictionary gives as one of the meanings "the intellectual side of civilisation"), in German the moral aspect is inseparable from the use of the word, except in the specialist writings of one or two philosophers.<sup>1</sup> Adelung, who, by his use of the word in the compound *Kulturgeschichte*, and, by his popular work on that subject, largely contributed to its acceptance as an equivalent for the English "civilisation," defines the word as meaning "the ennobling or refining of the whole spiritual and bodily powers of a person or a people, so that the word connotes not only the enlightenment (*Aufklärung*) and ennobling of the understanding by delivering it from prejudices, but also the polishing (*Politur*) and refinement of the moral qualities (*der Sitten*)."<sup>2</sup> At the beginning of the nineteenth century the word was used, in Germany, and it still is used, in the following figurative senses :

(I) To signify "state of civilisation" in its widest sense, without any implication of such a state being good or bad. It is this sense that it carries in the widely used expression *Kulturgeschichte*. The 1897 edition of Meyer's *Konversationslexikon* states that *Kulturgeschichte* requires an "investigation of the whole social life, including dwellings, hygiene, clothing, furniture and utensils, way of living, food, morals and customs, legal conceptions, beliefs and superstitions, at different periods." The distinguished English ethnologist, E. B. Tylor, in his book on *Primitive Culture*, uses the word exactly with such connotation as a synonym for "civilisation in its

<sup>1</sup> For example, H. Rickert, in his *Kulturwissenschaft und Naturwissenschaft* (1910), see pp. 26, 27 and *passim*.

## EXAMINATIONS—NOTE D

widest sense."<sup>1</sup> It seems clear that Tylor was influenced by the German use of the word "Kultur," and so gave to the English word "culture" a sense which it did not previously possess, and still only possesses in English as a term of ethnology.

(2) To mean "civilisation as opposed to barbarism," the "Kultur" of people who are regarded as being on a high plane. Goethe, in a remarkable passage in the *Conversations with Eckermann* (third part, March 14th, 1830), contrasts *Kultur und Barbarei*.<sup>2</sup> The sense of civilisation as opposed to barbarism is also well illustrated by the following passage from Schiller. Speaking of the influence of Italy on France in the Renaissance of the sixteenth century, he says: "The dawn of *Cultur* appeared. France was already hastening to welcome her civilisation" (*i.e.*, after the Dark Ages).<sup>3</sup>

From this limited sense of civilisation were de-

<sup>1</sup> See Preface to *Primitive Culture*, 4th edition. The first edition appeared in 1871.

<sup>2</sup> But a little later in the same passage Goethe uses it in the first and more general sense, and says: "There is something singular about national hatred; one always finds it strongest and most violent at the lowest stages of civilisation (*Kultur*)." Goethe is explaining to Eckermann why he declined to write songs of hate against France, and asks how he could write songs of hate without hating, and how could he, to whom culture and barbarism were the only things that mattered, hate the French, who were among the most cultivated of people, and to whom his own culture (here he uses the word *Bildung*, see footnote on p. 99 below) was so largely due.

<sup>3</sup> "*Die Morgenröthe der Cultur erschien. Schon eilte Frankreich mit schnellen Schritten seiner Civilisirung entgegen*" (*Geschichte der französischen Unruhen*, etc., K. Goedeke's edition of Schiller's Works, 1867, vol. ix. p. 294).

## EXAMINATIONS—NOTE D.

veloped three more specialised uses of the word, set out below.

- (3) Moral elevation. Goethe writes in his *Maximen und Reflexionen* (Anhang, Cotta edition, vol. iv. p. 230): "Piety (*Frömmigkeit*) is not an end in itself, but a means by which to attain through the purest spiritual quiet to the highest moral elevation (*Cultur*)."  
We find the same use of the word in *Dichtung und Wahrheit*, Part II. Book ix. (Cotta edition, vol. xxiii. p. 188). Goethe also sometimes uses the phrase "*sittliche Kultur*" (see, e.g., *Dichtung und Wahrheit*, Part II., Book vii., *loc. cit.*, p. 9).

(4) Individual refinement and sensitiveness, combined with knowledge, in the intellectual and æsthetic spheres. This is approximately the sense in which I have used "culture" in the text. It is a rare sense in German. An example of its use by Goethe is given in Note C, p. 94, above.

(5) External polish, relating to dress, manners, etc. It is in this sense that Goethe uses it in *Faust* (Part I. lines 2495-6), where Mephistopheles explains to the Witch his reason for concealing his horns and tail as out of fashion.

" Auch die Cultur, die alle Welt bedeckt,  
Hat auf den Teufel sich erstreckt."

which may be freely rendered :

" Of culture too, which all good folks now wear,  
The Devil himself has got a share."

With a word of such wide connotations individual differences of usage by different authors would in any language be inevitable, but the central fact in the history of the word "Kultur" in Germany is that

## EXAMINATIONS—NOTE

the most frequent and common sense in which the word is used is that of civilisation as opposed to barbarism, and with stress laid on the moral side of civilisation.<sup>1</sup> Absurd and ridiculous as it may seem, it is quite clear that it is because of this sense that the Pan-Germans in their own country have been able to use it as a signal and a watchword almost comparable to the Liberty, Equality, and Fraternity of the wars of the French Revolution. To us it now stands for the massacre and enslavement of Belgium, Serbia, Armenia, the repudiation of all international

<sup>1</sup> It has been stated that the proper translation of "culture" in the English sense is *Bildung*, and not *Kultur*. *Bildung* is a synonym of the fourth sense of *Kultur*, which is rare in German, and to which it corresponds almost exactly. Thus F. Paulsen says of *Bildung* that it consists not merely in the possession of things known (*Kenntnisse*) but in the possession of the living forces of active knowledge (*des Erkennens*) and of action in which the inner living individual (*die innere Lebensform*) takes part (Article on "Das Moderne Bildungswesen" in P. Hinneberg's *Die Kultur der Gegenwart* (1906). Goethe, however, seems to limit the word *Bildung* rather to a receptive assimilation of knowledge than an active one.

It has also been stated that *Zivilisation* in German means only material civilisation. That this is not strictly accurate is shown by the quotation from Schiller (on p. 97 above). Harnack, the theologian, in a passage written in 1909, says that England's "*Zivilisation*" is, taken as a whole, still superior to that of Germany, when, from the argument, it is quite clear that he means something much wider than material civilisation (*Aus Wissenschaft und Leben*, 1911, p. 200). On the other hand, W. Muench regards *Kultur* as consisting of "internal *Kultur*," often called *Kultur*, and "external *Kultur*," called "*Zivilisation*," the two being the inseparable wings of a "single front" (*Aus Welt und Schule*, 1904).

## EXAMINATIONS—NOTE) D .

law, the atrocities of Louvain and Rheims. The moral degradation of Germany has degraded the word itself. The degradation has not taken place without a long struggle, and the uses and definitions of the word *Kultur* by men of different political schools, Liberal and Socialist on the one side, and Pan-Germanists on the other, are historically significant of that struggle. Harnack contrasts ironically the *Kultur* of the present day, with its *arrière-pensées*<sup>1</sup> and compromises and concealment of its original source, with the spirit of scientific investigation (*Wissenschaft*) which knows no compromises and no *arrière-pensées*, and suggests that, if things had to be settled by contemporary authority, *Kultur* would be entrusted to "highly trained journalists;" while *Wissenschaft* would be shut up in a monastery.<sup>2</sup>

Paul Michaelis, a liberal journalist advocating a closer union between Germany and England in his book on Prussian Culture,<sup>3</sup> actually introduces into his definition of *Kultur* the phrase "the will for Peace." But Bernhardi, in his book on the Future of Germany, published in 1913,<sup>4</sup> demonstrates that an essential part of *Kultur* is the will for War, his

<sup>1</sup> Harnack uses the word *Rücksicht*; I know of no English equivalent for the word in this particular sense.

<sup>2</sup> *Reden und Aufsätze*, 1904, ii., p. 371; the passage itself dates from 1905. Harnack, alas, signed the Chauvinist manifesto of the German professors in 1914 to which the noble and moderate reply of the French Universities forms so striking a contrast.

<sup>3</sup> *Von Bismarck bis Bethmann. Die Politik und Kultur Grosspreussens*, 1911, p. 271.

<sup>4</sup> *Unsere Zukunft*, see pp. 40-43. The book has not, so far as I am aware, been translated into English.

## EXAMINATIONS—NOTE D

demonstration being based on the theory that "German *Kultur* is the most necessary factor in human progress," and that the acquisition of colonies by Germany is essential for the propagation of German *Kultur*. "*Kultur*" is the Red Cross flag under which the German pirates have sailed to the attack on humanity.

The object of this note has been partly to clear up some current misconceptions, but mainly to show that our fine word "culture," which has no synonym in English, ought not to be driven from our language by the degradation of *Kultur*. I hope to deal with the question more fully elsewhere.



## NOTE E

### PROF. EDGEWORTH'S INVESTIGATIONS ON THE STATISTICS OF EXAMINATIONS

1. THE only published statistical investigations of importance on the results of examinations of which I am aware, are contained in three memoirs by Prof. F. Y. Edgeworth, of the University of Oxford; (i) The Statistics of Examinations, *Journal of the Royal Statistical Society*, vol. li., pp. 599-635 (1888), (ii) The Element of Chance in Competitive Examinations, *id.*, vol. liii., pp. 460-75 and 644-63 (1890), and (iii) On Problems in Probabilities (*Philosophical Magazine* for August, 1890).<sup>1</sup>

2. *Definition of Standard or True Mark and general Method of Investigation.*—In comparing the marks of different examiners with one another, it is necessary to

<sup>1</sup> These papers also contain references to other works by the same distinguished author, including a treatise on the art of measurement, entitled *Metrelike* (London: Temple Co.). An elaborate account of the mathematical theory of Probability is given in Professor Edgeworth's article on that subject in the eleventh edition of the *Encyclopædia Britannica*. The first two papers, which are summarised below, are intended for the general reader; the third contains a more technical treatment of problems dealt with in the first two. The present writer has no expert knowledge of statistics, and is greatly indebted to Professor Edgeworth for having revised this summary.

## EXAMINATIONS—NOTE E

assume that there is some standard, or "true" mark. In the second paper, Prof. E. postulates "that the true or standard mark of any piece of work is the average of the marks given by a large number of competent examiners equally proficient in the subject and instructed as to the character and purpose of the examination."

Prof. E. clearly states this as a postulate, not as an axiom; but it is a convenient postulate, and is probably generally acceptable. The general method adopted by Prof. E. is to ascertain what this standard is in a number of typical cases, and to investigate the divergency from this standard, or *error*, of the marks allotted by individual examiners, and to deduce general laws of such errors, with the object of ascertaining the probability that an error of assigned extent will be committed by a competent examiner in marking a single piece of work of this kind; and from this to deduce the probable errors in the more complex cases which present themselves in practice.<sup>2</sup>

<sup>1</sup> This is a development of the corresponding statement in the first paper.

<sup>2</sup> Even at the outset there are at least two practical difficulties to be got over: (a) the difficulty, involving as a rule expense, of obtaining a large number of "equally competent" examiners to look over and mark the same piece of work; and (b) the difficulty of obtaining any definition of the "purpose" of the examination in question. As I have pointed out in the text, this absence of definition of purpose constitutes, in my judgment, one of the great weaknesses of our present system, and one which necessarily leads to far greater divergencies of marking than would otherwise exist. Prof. E., in his second paper, expressly states that his subject is "not the philosophy of examination, but the statistics of marks," and does not attempt therefore to go behind the marks allotted.

## EXAMINATIONS—NOTE 'E

3. *Mutual compensation of independent errors.*—*Effect on aggregates of marks.*—Prof. E. points out that, according to the Calculus of Probabilities, if a certain quantity is liable to a probable error of 1 per cent. in its measurement, the probable error on the aggregate of the sum of a hundred such quantities is not 1 per cent., but only  $\frac{1}{10}$  per cent., that is 1 per cent. divided by the square root of 100, "in virtue of the mutual compensation of independent errors." The statistical theorem of which this is an example is much used throughout his investigations.<sup>1</sup> (See also § 24, p. 130.)

4. The above considerations are common to the two papers, of which I now submit separate summaries, necessarily brief and imperfect, but bringing out certain salient facts.

<sup>1</sup> The whole investigation is based on the Mathematical Theory of Probabilities. The following definitions and statements will be useful to the reader:—

(1) The probable error of a series of observations is that divergence from their mean on either side . . . within which exactly half the observations lie. An approximation to the probable error for a given series of observations is obtained by arranging all the observations in order of magnitude; marking the magnitude, say,  $\alpha$ , above which 25 per cent. of the observations lie, and the magnitude, say  $\beta$ , below which 25 per cent. lie. Half the difference between  $\alpha$  and  $\beta$  is the probable error.

A useful way of illustrating this is to say that if one observation is chosen at random out of a group it is as likely as not that it will not lie further from the average than the probable error. (A. C. Bowley's *Elements of Statistics*, second edition, 1902, pp. 281-2.)

(2) If  $n$  marks (or other magnitudes) are arranged in an ascending order of magnitude the mean of the  $\frac{n^{\text{th}}}{4}$  mark

# EXAMINATIONS—NOTE

## PAPER I<sup>1</sup>

5. *Choice of an Ideal Typical Examination for Investigation.*—In order to make the subject more readily intelligible, Prof. F. takes as a particular case

and the  $\left(\frac{n}{4} + 1\right)^{\text{th}}$  mark is called the lower quartile; the mean of the  $\left(\frac{n}{2}\right)^{\text{th}}$  and the  $\left(\frac{n}{2} + 1\right)^{\text{th}}$  marks is called the mean of the set; and the mean of the  $\left(\frac{3n}{4}\right)^{\text{th}}$  and  $\left(\frac{3n}{4} + 1\right)^{\text{th}}$  marks is called the upper quartile. The probable error is thus half the difference between the upper and lower quartiles.

It will be seen that the "possible error" is always considerably in excess of the probable error.

(3) The "mean error" of a series of  $n$  observations of which the individual errors or differences from the mean value are  $e_1, e_2, e_3 \dots e_n$ , is the square root of the sum of the squares of those errors divided by their number, i.e.,

$$\sqrt{e_1^2 + e_2^2 + e_3^2 \dots + e_n^2}$$

For aggregates fluctuating according to the ordinary "law of error" the "mean error," so defined, is about 1.25 times the average error, i.e.  $\frac{e_1 + e_2 + e_3 \dots + e_n}{n}$ , each  $e$  being taken positively, or in absolute quantity, as the phrase is, without respect to sign.

The "frequency curve" of a series of magnitudes (such as a series of marks or errors in marking) is a curve of which the abscissæ represent the magnitudes and the ordinates the comparative frequencies with which the corresponding magnitudes occur in the series. The frequency curves for errors in marking (and especially for errors on aggregates of marks, see § 17 below) appear in general to belong to the special type called *the curve of error*, of which Prof. Edgeworth likens the shape to the longitudinal section of a *gendarme's hat*.

(2) and (3) are derived from Prof. E.'s papers.

<sup>1</sup> For Summary of Paper II see p. 118 below.

## EXAMINATIONS—NOTE E

a typical examination in which candidates are arranged according to the aggregate of marks which they have received on ten different papers, for each of which 300 marks are allotted, and assumes "that those who, without failing in any subject, have not obtained in all the subjects more than 2,000 marks, are placed in alphabetical order in the category of 'Pass'; and those who have obtained more than 2,000 are placed in the category of Honours and arranged in 'order of merit' according to the number of marks received." The paper investigates the error to which these aggregates of marks are liable; and thence proceeds to deduce the degree of certainty incident to the distinctions grounded on these marks.

No attempt is made to estimate the chances introduced by the possibility of a candidate being "out of sorts" on the day of examination, or of the paper being specially adapted to the accidents of his taste or reading; "it is assumed, as in the case of our public examinations may be safely assumed, that the paper set in each case is an ideally fair one."

6. *Errors due to Minimum sensible.*—Prof. E. next points out that there must be a *minimum sensible* in our perception of excellence, of which the magnitude will vary with the subject, being least in Mathematics and most in subjects where marking depends greatly on "taste."

Where subjects are for the most part of the simpler sort, Prof. E. estimates the "probable error" of the mark for each answer at 5 per cent.; in other words, when an examiner is putting down 20 (the maximum being larger) he very generally hesitates between 19 and 21 as equally representing to him the worth

## EXAMINATIONS—NOTE E

of the answer. Prof. E. then supposes that the average number of questions attempted is 80, and deduces that the "probable error" under this head, is at least  $\frac{5}{\sqrt{80}}$  per cent. or  $\frac{1}{2}$  per cent. (see § 3 above).<sup>1</sup>

Thus if marks between 2,000 and 2,050 had been assigned to a candidate, there is an uncertainty on this account represented by a "probable error" of 10. If the aggregate of marks follows what is often called the *normal* law (or frequency curve), this corresponds to a "possible error" of 20, the odds against such an error being only 6 to 1, and a "very improbable error" of 40, against which the odds are 250 to 1. Even on the most limited view of the influence of chance a considerable uncertainty arises as to the true position of candidates in the Honours list. It is to be remembered that this error is altogether unavoidable.

**7. Errors due to Personal Equation of Examiners.**  
*Variations of two kinds.*—A much more serious cause of inaccuracy is the "personal equation" of the examiners. Two species of variation are to be distinguished under this head, (1) variations relating to particular answers, and (2) variations relating to the whole scale of marking.

<sup>1</sup> Thus in a pass examination in which pass or failure depends on the result of two papers, and say sixteen answers in all, the probable error would be  $\frac{5}{\sqrt{16}}$ , or 1.25 per cent.

I cannot help regretting that the author did not deal more completely with the simple case in which "pass" and "failure" depend on the results obtained in separate subjects, independently of the aggregate of marks obtained in all subjects.

## EXAMINATIONS—NOTE E

Prof. E. states that his data (which are not quoted in full) show that the observed mean difference per cent. between two (numerous) sets of marks is 20. Taking into account an error of 5 per cent. due to the existence of a *minimum sensibile*, he finds the probable error due to the first cause to be 10 per cent. ("corresponding to a just possible error of 30 per cent."). In the typical examination dealt with the probable error will be  $\frac{10}{\sqrt{80}}$  i.e. 1.1 per cent.<sup>1</sup>

Prof. E. states that an experienced Oxford examiner tells him that, in general, where examiners are careful and experienced, "marks run very close; within (say) 10 in translations (full marks being 100), within a class, or, better half a class, in subjects and essays (there being four classes, and the mark given being one of the letters  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ , qualified in some cases by a + or a — to indicate a doubtful or intermediate position").

Prof. E. gives, as an illustration of the second kind of variation, the difference of scale, a case of an examination in English in which he and a colleague "had agreed beforehand as to the worth of the different questions, and as to the degree of proficiency to be expected of the candidates," yet the average of his colleague's marks came out 13 per cent. higher than his own, the former being 165 (out of a maximum of 300) the latter 145.<sup>2</sup> Thus, if the examinations had been conducted on one occasion by the first examiner

<sup>1</sup> If, as before, we take only sixteen answers, the probable error under this head will be 2.5 per cent.

<sup>2</sup> The percentage refers to the mean of the two averages, viz. 155. (Footnote by Prof. E.)

## EXAMINATIONS—NOTE E

and on another by the second, there would have been a difference of 13 per cent. in the average marks allotted.

Prof. E. supposes "that in so well organised a system as that of our public examinations, care is taken to avoid this disturbance," and that there is a "tradition as to the relative value of each subject to be preserved by the slowly changing staff of chief examiners."<sup>1</sup>

Still he says, "in spite of conscious efforts to maintain constancy, it would not be unreasonable, I think, to assume a probable error of 4 or 5 per cent. due to the factors under consideration. Thus, supposing there was no real difference between the papers in Mechanics of this year and the preceding, that the real average, so to speak, of both sets of papers was 150, yet in the sortition of examiners it is likely as not that the average may be pushed up to 156 or down to 144; and it is possible that it may deviate much further from the ideal mean."

Prof. E. adds in a footnote that, in making this estimate, he stands astonished at his own moderation, "for there is no doubt that at many of our public examinations the fluctuations in the scale are enormous." He quotes Dr. Venn, "an experienced examiner, who is also one of the highest living authorities on statistics," as saying, "I have frequently raised or depressed my own marks (or my colleagues') by as much as 25 per cent. all through in order to bring them into general harmony."

<sup>1</sup> I presume that Prof. E. refers here to such examinations as Class I. Civil Service Examinations, where not only candidates, but "subjects" compete.



## EXAMINATIONS—NOTE E

Adopting, however, the probable error of 4·5 per cent. in respect of personal equation, the probable error on 10 papers will be  $\frac{4\cdot5}{\sqrt{10}}$  or about 1·5 per cent.<sup>1</sup>

An increase or diminution in the scale of marking in a given subject will not affect the relative order of candidates in that subject; but it will obviously affect the relative order of the aggregates of marks obtained by candidates of whom only some take this subject, and therefore their relative order in the Honours list.<sup>2</sup>

Prof. E., by reasoning which he does not give in full, comes to the conclusion that the additional probable error on the aggregates due to difference of scale is only from  $\frac{1}{3}$  to  $\frac{1}{2}$  per cent. "and may safely be neglected."<sup>3</sup>

8. *Errors due to Correction at great speed—or the "negligence of haste."*—He next investigates the subject of differences in marking due to difference of speed; and finds that the *minimum sensibile* for papers

<sup>1</sup> If there are only two papers it will be  $\frac{4\cdot5}{\sqrt{2}}$  or about 3 per cent.

<sup>2</sup> Prof. E. does not deal with the question of how far the pass in each subject is affected. Suppose the pass mark is 33 per cent., the probable error on the 33 marks out of 100 is 4. As I have said elsewhere, I find it difficult to attach any definite meaning to such a pass-mark as 33 per cent. on capacity tests, although it has a definite meaning in memory tests (see p. 63). It ought to be said that in many examinations, probably most, border-line cases receive special consideration which must reduce the probable error near the border-line, though it clearly cannot be eliminated, since the judgment of different examiners is bound to vary.

<sup>3</sup> As will be seen from § 10 below, Prof. E. finally decided at a later stage not to neglect this error.

## EXAMINATIONS—NOTE

corrected at top speed becomes considerable. Taking this into account, he finds a probable error due to "negligence of haste" of about 25 per cent. If there are 80 items affected by this error, the probable error of the aggregate is  $\frac{25}{\sqrt{80}}$  about 2·8 per cent. But the author assumes that the results in public examinations are, when Honours are concerned, subject to "comparatively leisurely revision and re-examination by certain chief examiners," and therefore does not take the item into account in dealing with the typical examination under consideration. He regards rapid examining as "safe enough for the purpose of a Pass examination."

9. *Errors due to Fatigue of Examiners.*—Prof. E. next investigates the error of marking due to the fatigue of examiners, which might be avoided if more time could be allowed. He thinks this may amount to from 1 to 2 per cent. on each paper; allowing 1·5 per cent. and 10 papers, the probable error is  $\frac{1.5}{\sqrt{10}}$  or say  $\frac{1}{2}$  per cent. on the aggregate marks for the examination.

10. *Summary of Errors due to marking in Typical Examination.*—For the typical examination Prof. E. sums up his results as follows:—

	Percentage (probable) Error of Aggregate.
Due to existence of a <i>minimum sensible</i> . . . . .	$\frac{1}{2}$
Due to the examiner's idiosyncrasies :	
(1) as regards particular questions . . . . .	1 $\frac{1}{2}$
(2) as regards scale . . . . .	1 $\frac{1}{2}$
Due to the "negligence of examiners" as regards scale . . . . .	$\frac{1}{2}$

The total probable error expressed as percentage of

## EXAMINATIONS—NOTE 'E

the aggregates of marks is the square root of the sum of the squares of the separate items, or rather more than 2 per cent.

"For instance, if a candidate obtains an aggregate of 1,965 marks, thus falling below the Honour line by 35 marks, the odds that he ought by right to be in the Honours Class are considerable, say one in four. Even if he has obtained only 1,800 marks, it is by no means certain that he is rightly placed among the rank and file of Pass candidates."

The same doubt exists as to the rightful position of those who are the same distance above the Honour line. "To such an extent are the Honours shuffled by the fortune of examinations!"

11. *Comparison of foregoing Results with certain Statistics of Cambridge Classical Tripos.*—From an investigation of the marks of a Cambridge Classical Tripos examination, Prof. E. comes to the conclusion that the probable error on the mark of a single paper is not less than 9 per cent., and on the aggregate mark of 9 or 10 papers, 3 per cent.

Prof. E. gives an interesting table, too long to quote here, of the aggregates of marks for ten papers assigned to fifteen candidates by two examiners in each case. The greatest difference between two aggregates is 59, the two aggregates being 472 and 413; the least difference is 8, the two aggregates being 632 and 624. Calling the difference between the two aggregates in each case and their mean the "deviation," we find the average deviation or "error" per cent. (of the mean marks in each case) to be 3.5. The

## EXAMINATIONS—NOTE II

"probable error" is .84 of the average error or about 3 per cent. (See also § 19, p. 125).

12. *Error due to possible non-representative Character of Answers of Candidates—Total Error of Marks in Typical Examination.*—Prof. E. touches on the difficult problem of evaluating the error committed in taking the answers of the candidate as representative of his proficiency (in other words, using the language of the text, what is the probability of the sample which he has given of his proficiency being a fair sample). He estimates the variation of marks, from answer to answer to be adequately represented by a probable error of 25 or 30 per cent. ; and on this basis calculates the probable error on the aggregate of marks for typical examination to be about 2 per cent.

Cumulating this error with the account already furnished, "we find for the aggregate of marks, considered as a test of the candidate's real proficiency, a probable error of about 3 per cent., corresponding to a possible error of more than 9 per cent. It is to be remembered that this estimate leaves out of account many elements of uncertainty which are too subtle for the Calculus of Probabilities to handle; such as the variation of the candidate's spirits, or the circumstance that the questions are not fairly representative of the subject. . . . It is to be remembered, also, that the error in appreciating the answers is reckoned from the supposed ideal mean which would be presented if a large number of competent sets of critics were to examine the work. If it is asked, what difference is to be expected between the results of any two sets of examiners,

## EXAMINATIONS—NOTE E

the estimate of probable error must be somewhat increased."

13. *In how many Cases may award of Honours in Typical Examination be due to Chance?*—Prof. E. next touches on the question (1) in how many cases the obtaining or not obtaining Honours may be regarded as accidental, and for this purpose makes the assumption, which he thinks legitimate, that out of every 1,000 candidates, there are among those who have passed, at least 120 who have obtained marks between 1,800 and 2,000. On this basis he calculates that the average number of candidates who are placed in the wrong category owing to chance is about 30, although in one year it might be only 20 and in another rise to 40.

14. *Recommendations for Reform:* (1) *Honours Lists.* (2) *Working of Assistant Examiners.* (3) *Question of making Examination-marks conform to a given Frequency Curve.*—To mitigate the defects arising necessarily from chance, Prof. E. makes certain recommendations "based on purely statistical reasons," which he is quite prepared to find may be overborne for purely practical reasons.

(1) *Arrangements of Honour's Lists.*—He suggests that "the arrangement of an Honours List in an ostensible order of merit should be abandoned when the distinctions thus drawn do not correspond with any degree of probability to real differences. The extent of each homogeneous alphabetically arranged class should be considerably greater than the probable error incident to the aggregate of marks upon which the arrangement is based. Thus, in the example chosen, if an aggregate of marks between 2,000 and

## EXAMINATIONS—NOTE E

2,100 is liable to a probable error of 50, then all or most of the candidates between these limits might be placed in one homogeneous class.<sup>1</sup> •

(2) *Question of Assistant Examiners.*—He suggests that the functions of Assistant Examiners requires special consideration, and says there appear to be two distinct theories, (a) and (b), with regard to the arrangement in respect of Honours candidates who [are reported by the assistant examiners to] have passed; and that the two theories may admit of intermediate shades and practical compromises.

(a) According to theory (a) the function of assistants is merely to select provisionally, for careful re-examination *ab initio* by the chief examiners, candidates who may be deserving of Honours.

In this case, the papers selected for re-examination should be defined, not by the Honour line of 2,000 marks, but by one considerably below it, due account being taken of the *ex hypothesi* rough marking of the assistants being affected by a very large probable error. Prof. E. thinks that, according to this view, in order to avoid mistakes it would be necessary for the chief examiners to re-examine completely and *de novo* the papers of about half the candidates who have passed.

<sup>1</sup> Professor Arthur Schuster, on the other hand (*University Review*, May 1905), has pointed out that the differentiation into classes without any indication of order of merit, may make an unjust distinction between the last in one class and the first in the next. The creation of a large number of classes in a typical examination such as that postulated by Prof. E., with a class in alphabetical order for each multiple of 100 marks, might be a satisfactory compromise; it would not be necessary to make use of the word class, the candidates within each century would merely be bracketed.

## EXAMINATIONS—NOTE E

(b) According to this theory, the marks assigned by the assistant examiner "are final, subject to the inspection of the chief examiner, and the revision of exceptional cases." According to this principle, "or any intermediate plan in so far as it involves this principle, it may be expected that the sources of error which have been above described under the head of negligence<sup>1</sup> will become peculiarly operative."

Prof. E. at once mitigates this criticism by a footnote, in which he says: "What may be safely affirmed of any public examination is, on the one hand, that any method which has commended itself to many able and experienced examiners is fairly good; but that, on the other hand, it is not likely to be the best possible, if it has been adopted without reference to the theorems in Probabilities which evidently dominate the whole subject."

In order to avoid the error due to marking at high speed, Prof. E. next investigates the possibility of combining rapid dealing and accuracy by omitting the correction of half the answers in each subject and comes to the conclusion that the probable error incident to the aggregate of marks obtained by doubling the marks so allotted (the correction being leisurely) is only 1·4 times the probable error (and not twice the probable error) of the aggregate obtained

<sup>1</sup> I think the word "negligence," though used by Prof. E., implies a moral stricture on the individual examiner or assistant concerned which is not intended by the author, who points out, in dealing with this matter, that "the number of the staff in relation to the quantity of work to be done, and the time for doing it, is sometimes such as to necessitate an extraordinarily rapid rate of examining." Prof. E. tells me he concurs.

## EXAMINATIONS—NOTE E

by correcting all the answers, the correction being leisurely.

Thus if "the total aggregate of marks, considered as representing the real proficiency of the candidate is liable to an error of 2 or 3 per cent. [see § 10, above] the error incident to an examination of half the answers is only 3/5 or 4/5, the number of answers being large, and the half selected at random."

Prof. E. suggests, as a simple method of diminishing the time taken over correction which would avoid hasty correction is to increase the number of assistant examiners, and further that the answers of the candidates should be separable, the answers to (say) the first two questions of all the candidates being looked over by examiner A, the answers to the second two by examiner B, and so on. "By this arrangement the influence of the personal equation, introduced by a plurality of examiners," would be so distributed as to become insensible.<sup>1</sup>

### (3) *Question of making Examination-marks conform*

<sup>1</sup> This arrangement is adopted in certain examinations. It has the further advantage that the examiner, who necessarily marks slowly at first, and increases speed as he becomes more familiar with the difficulties of the paper and the problems involved, attains the maximum speed consistent with care more quickly than if he has to deal with, say, four or five times as many different questions. The disadvantage is that he cannot judge of any candidate's work as a whole. It sometimes appears from the correction of a whole paper, in a way that correction of a part would not reveal, that the candidate is making correct use of terms which he does not in the least understand. Hence it would appear, as Prof. E. points out, that the principle would require "to be tempered with practical sagacity and knowledge of details" before it could be immediately applicable.



## EXAMINATIONS—NOTE E

to a given Frequency Curve.—Prof. E. suggests that it may be possible, in each subject, to determine what ought to be the proper average for a set of marks of different candidates at a given examination and the proper “dispersion” of the marks about this average; and he shows how another set of marks may be reduced so as to conform to this arrangement. The suggestion is repeated in the second paper in respect of Government examinations (see § 25 below).<sup>1</sup>

### PAPER II

15. *The Differences of marking between different Examiners, in respect of one Paper, are similar to the corresponding Differences in respect of another Paper in the same Subject.*—Comparison in the case of different subjects (*Geometry and History*).—After certain preliminary matter (see § 1 above) the author discusses the general question of laws of error attaching to different kinds of judgments or estimates, relating

<sup>1</sup> In revising this abstract, Prof. E. draws attention to the great importance of paying attention to “dispersion” when marks are combined, and points out that if the award of a prize is decided by the sum of marks obtained in Latin and Greek, and if, while the average obtained in the two subjects is the same, the “dispersion” is much greater in Greek, the Latin marks will tend to become inoperative. He illustrates the argument by curves which cannot be reproduced here; but it is easy to see that if the marks of, say, the first half-dozen candidates lie close together in Latin, but are widely dispersed in Greek, the Greek marks would decide the matter. This result would be unfair if it were due only to the difference of idiosyncrasy of the Latin examiners and the Greek examiners in their method of assigning marks. In such a case verbal discussion of the results is essential for a fair award.

## EXAMINATIONS—NOTE E

to things other than examination answers, e.g. weights and heights. He discusses the case of marks awarded by sixteen different competent examiners to the work in mathematics of a pupil in the sixth form of the North London Collegiate School, and finds the probable error deduced from the sixteen observations to be 4·5 per cent. "For the eight first in the order of entry it is 3·5; for the second batch it is 4·5. Accordingly, it is not a leap in the dark when we infer that what is true of one set of sixteen examiners marking that paper would be approximately true of another set of equally competent examiners." He finds, further, that the law of dispersion for two *different* papers corrected by the same examiners is also approximately the same. Although he expected the variation in the marking of history papers to be greater than in the case of mathematics, the limited experience afforded by three sets of marks assigned by nine different examiners to three history papers showed no material difference between the two cases.

### 16. *Marking of piece of Latin Composition by twenty-eight different Examiners.*

I think it well to quote the following passage verbatim:—"The experiment which I shall next adduce is *in alia materia*. It relates to a piece of Latin prose which the Editor of the *Journal of Education* kindly allowed me to insert in a recent number of that periodical, with a preface which it may be well here to reproduce in part:—

I propose, through the medium of the *Journal of Education*, to invite any competent person to assign a mark to the subjoined piece of Latin

## EXAMINATIONS—NOTE 'E

prose, upon the supposition that he is marking the work of a candidate for the India Civil Service. Let it be distinctly understood that in giving his mark the examiner is not to look to, or wish to illustrate, his own ideal of classical elegance, nor yet the degree of proficiency which may be current in the school or other institution with which he may be connected. Let him imagine that he has been appointed examiner in Latin for the India Civil Service, and let him give his mark having regard only to what may be expected from a candidate for that prize. Let 100 be the maximum attainable by any candidate.

“To avoid accidental divergence as much as possible, to perform the experiment under the most favourable conditions, I would suggest that the examiners should consist of a pretty homogeneous class—of much the same class as those who actually conduct our public examinations. To be more definite, I would invite to take part in this experiment only those who have taken high honours in classics at one of the universities, or classical masters of the sixth form in a public school. All such are earnestly invited to examine the accompanying piece with as much care as if they really were exercising the function of a public examiner; and to send to the editor their verdict, guaranteed by their name and status, which, it need hardly be added, it is not intended to publish. It is desirable that the examiners should assign their respective marks independently, and without mutual conference.’

## EXAMINATIONS—NOTE F

"In response to this appeal twenty-eight highly competent examiners were so kind as to mark the piece of Latin prose. The grouping of these twenty-eight independent marks may be expressed by the accompanying scheme of figures, with the interpretation of which the reader will by this time be familiar<sup>1</sup>.

II	V	VII	VII	V	II
63'5	73'5	78'5	83'5	93'5	

"The actual marks are:—45, 59, 67, 67'5, 70, 70, 72'5, 75, 75, 75, 75, 75, 75, 77, 80, 80, 80, 80, 80, 82, 82, 85, 85, 87'5, 88, 90, 100, 100.

"It will be seen that 78'5 is the central value, the presumably correct mark<sup>2</sup>; the probable error on either side of the correct mark is 5. It may excite surprise that this error is not very much greater than what we observed in the case of the geometry and history, which, without offence, may be described as not requiring on the part of the examiner such an exercise of literary taste as the Latin prose. But when we consider the extreme as well as the central portions of the Latin prose group, we shall be justified in attributing to the marking of the more advanced work a larger element of chance. It is interesting to observe that the two highest marks, 100, are (each) more than double the lowest mark, 45. An occurrence of this kind is not at all uncommon in the

<sup>1</sup> The Roman numerals express the number of awards of marks of which the amounts lie between the Arabic figures on either side of them. Thus there were two marks below 63'5; five between 63'5 and 73'5; seven between 73'5 and 78'5, and so on.

<sup>2</sup> The arithmetic mean is 78 (Prof. E.'s footnote).

## EXAMINATIONS—NOTE 'E

marking of advanced work, where there is room for great diversity of taste in the examiners.<sup>1</sup>

16. Prof. E. deduces (with the help of the results set forth in § 3 above) that "in an examination of one subject admitting as much diversity of judgment as the Latin prose . . . a difference of 25 per cent. between the marks of two candidates may be accidental and liable to reversal. Suppose, for instance, that, under one examiner, candidate A. obtains 90, and candidate B. 70, that is, a difference of 25 per cent. of the average between the two marks, namely

<sup>1</sup> No doubt the extremer divergencies are partly due to a difference, not of taste, but of *scale*. In the case of English composition and the speculative sciences, I have met with equally great discrepancies in the marks given by two examiners to a particular candidate, without any corresponding difference in the average of their marks given to a whole set of candidates. (Footnote by Prof. E.)

Prof. E. now adds the following interesting note in regard to this experiment:—"As to the Latin prose, I ought perhaps to explain that in one respect it was hardly typical. Having written it as well as I could, I deliberately dashed it with an elementary error, the use of the preposition *penes* with the *ablative*, which, as far as I know, never occurs. This schoolboy error, no doubt, as I intended, utterly puzzled the judges. One broad-minded man, now the Provost of a distinguished College, said that such a mistake in such a paper might be treated as a mere *lapsus plumæ*, and he gave it full marks. Another, I think the one who marked 45, thought the mistake was unpardonable."

According to my own experience, Professor E.'s intentional mistake increases rather than diminishes the value of his experiment. A friend of mine who was *proxime accessit* for either the Hertford or Ireland, was told by the examiners that except for one thing he would have been awarded the prize; he had made one unpardonable mistake in a quantity in a Latin verse.

## EXAMINATIONS—NOTE E

80 . . . It is not a very improbable event—say about as unlikely as throwing heads at pitch-farthing six times running [say 1 to 100]—that under another examiner the order of the candidates might be reversed, B coming out above, though probably not high above, his rival."

**17. Error of Aggregate of Marks formed by Addition of Totals for different Subjects.**—It is necessary to deal with aggregates of marks formed by the addition of totals for different subjects for which the laws of error are different; but Professor E. states that "When a compound mark is formed by the addition of several marks on different subjects, the manner of divergence, or law of error, appertaining to the compound is of one and the same general type, whatever the type of the particular laws appertaining to the several components. For instance, let the several subjects be geometry, history, Latin prose, etc., and let the law of error proper to each of these subjects be obtained by observations such as those which have been adduced in the preceding pages. Then the compound (mark in geometry + mark in history + mark in Latin prose + etc.) will vary with the change of examiners according to a law of error not in general identical with that of the components, but of a certain normal universal type which may be, and is often, called *the law of error*."

Prof. E. states that what is true of sums holds good for differences, and this result is used in the last paragraph of § 16 above.

**18. General Case of Distinction awarded for Aggregates of Marks in different Subjects—Use of Results obtained from two Markings of the same Work by the**

## EXAMINATIONS—NOTE E

*Examiner at different Times.*—Prof. E. now deals with the general case in which distinction is determined by the sum of marks in different subjects.

He suggests that if we had to depend on data yielded by marking of the same work by several examiners, we might despair of obtaining sufficient statistical material for the reasoning, but thinks it suffices to combine a few instances of this type with a great many instances of two examiners both marking several pieces of work.

Prof. E. investigated the results of examination by the same examiner (Mrs. Bryant, D.Sc.) of the same forty papers written by her pupils, on two occasions separated by a considerable interval of time. He does not give the result directly,<sup>1</sup> but states that if the same candidate were examined in ten papers by two sets of ten examiners, and if each couple of examiners "marked on the whole at the same level but with respect to particular papers disagreed with each other about as much as Mrs. Bryant with herself, then the difference between their marks would in the long run as often as not amount to 2·5 per cent. (of the average mark for the whole set of ten papers examined)."

Thus, if a candidate obtains 2·5 per cent. of the average mark for the whole set of candidates above the mark for the Honours line, it is as likely as not that under another set of examiners he would have come out *either* at least as high as *five* per cent. above the line, or at least as low as the line. It is with the latter alternative only that we are concerned. The

<sup>1</sup> Further details are given in the paper by Prof. E. in the *Philosophical Magazine* for August, 1890.

## EXAMINATIONS—NOTE F

probability of this is only half a half, that is, a quarter.<sup>1</sup>

If a man clears the Honours line by 10 per cent. of the average marks, the odds are 300 to 1 against a reversal by a different set of examiners; if he clears them by 9.5 per cent., the odds are 200 to 1 against reversal; if he clears them by 7 per cent. they are only 50 to 1.

In the case of the award of appointments it is, as a rule, not an Honours line that is fixed in advance, but the number of appointments awarded. Prof. E. assumes that the difference of marking between examiners will be not greater than that between the two markings by Mrs. Bryant of the same papers on different occasions, and applies this to the case of an examination for fifty clerkships in 1875, for which there were 172 candidates. He comes to the conclusion that there are only 17 candidates for whom the odds against being displaced if there were a change of examiners is as great as 50 to 1.

**19. Details in regard to a Classical Tripos Examination.**—Prof. E. discusses a series of three sets of marks given by three examiners to 75 papers in composition. The average of two of the examiners was about the same, but the average of the third examiner was about  $\frac{1}{8}$  less than this and was increased therefore by 20 per cent. for purposes of comparison.

<sup>1</sup> In other words it is equal odds whether he is within the limit of 5 per cent. above Honours, or outside; it is 4 to 1 against his being outside the limits below the lower line, and it is 4 to 1 against his being outside them above the upper line. If a large number of cases were taken, one half would fall within the limits, one quarter below them, and one quarter above them.



## EXAMINATIONS—NOTE E

The coefficient of divergence might then be expressed as a "probable error of 5 per cent. on the sum of 10 marks."

"The severer examiner is as often as not distanced by one of his colleagues to such an extent that one mark is double the other. Fourteen times out of ninety, one mark is three times as great as another given to the same work. Four times, the excess is in the ratio 5 to 1. And once the severer examiner exceeds his more lenient colleague in the ratio of 16 to 1. These graver discrepancies may no doubt be obviated by properly correcting the scale of the severer examiner. Still the effect of the difference in scale, and the efficiency of correction, may easily be exaggerated. It may surprise even experts to hear that, as between two examiners, A and B, who upon the whole are fairly well abreast, occasionally A is ahead of B, and B ahead of A to the extent of 50 per cent." (See also § 11.)

20. *Details in regard to Oxford Literæ Humaniores Examination in Philosophy.*—Prof. E. finds the uncertainty in the marking at the Cambridge Classical Tripos to be about the same as the uncertainty in the marking in Philosophy at the Oxford Final Honours in *Literæ Humaniores*, of which he gives some details. He points out that the fluctuation in the estimates of those who are in or near the first class is markedly less than for less successful candidates. "In fact," he says, "it is as small as any which I have observed in the case of subjects more elementary and precise"; a result which he regards as a striking tribute to the Oxford method of marking.

21. *Details of Examinations in English Composition*

## EXAMINATIONS—NOTE E

*and Speculative Science.*—Prof. E. next investigates the marks allotted by two different examiners to 400 pieces of English composition. The “mean error” (i.e. the square root of the sum of the squares of the 400 differences between the individual pairs of marks, divided by  $\sqrt{400}$ , is 67 marks.<sup>1</sup> If, instead of taking the whole 400 observations, they are divided into three batches of 133, the corresponding figures are 63, 67, and 72.

The author next points out that there was a difference of “scale” between the two sets of marks amounting to 10 per cent. of the average mark. Reducing the marks to the same scale the “mean error” is only reduced to 64 marks.

The average of the 800 marks for the whole set of papers then becomes 227; and the “mean error” 64 is about 28 per cent. of this average. If there were ten papers of this kind, the “mean error” for the ten would be  $\frac{28}{\sqrt{10}}$  per cent., and the “probable error” would be 6 per cent.

Thus, on the aggregate marks for the papers for each of which there was a “mean error” of 28 per cent., the error as likely as not to occur would be 6 per cent. of the average mark. In a public examination on speculative science “not guarded so carefully as the Oxford Examination in Philosophy,” the probable error for a set of ten papers amounted to 8 per cent.

**22. Summary of Previous Results.**—The author now sums up his results in the following table of the discrepancies as likely as not to occur between the sums

<sup>1</sup> Cf. footnote on pp. 104-5 above.

## EXAMINATIONS—NOTE E

of marks on ten papers assigned by two sets of ten examiners to the same papers. "The reader will recollect that each of these coefficients requires to be multiplied by at least 3'5 or 4 in order to obtain the height above the Honour line at which a candidate's position may be regarded as irreversible."

	Number of papers scrutinised.	Probable error calculated for a set of 10 such papers, expressed as percentage of the average aggregate mark.
High School, 6th Form, Geometry. (Marked twice by the same examiner at different times [see § 18])	82	2'5
High School, 6th Form, Geometry and History	160	3'0
Cambridge Classical Tripos, Composition and translation, History and Grammar (mixed) (see § 19)	480	4
Latin prose of the type set at the I.C.S. Examination (see § 16)	28	4'5
Oxford Literæ Humaniores, Philosophy only (see § 20)	460	4'5
Cambridge Classical Tripos, Composition only (see § 19)	225	5
English Composition (see § 21)	800	6
Speculative Science (see § 21)	140	8

**23. Question of Measure for Element of Chance in any Public Examination.**—Prof. E., "inclining to the side of moderation," adopts as the probable error for marks at public examinations the lowest based on the comparison of results of different examiners, viz. 3 per cent. of the average mark obtained at the examination.

## EXAMINATIONS—NOTE E

"Adopting this datum," he writes, "I am able to answer the following question with respect to any examination of which the marks are published: How many of the candidates placed are quite safe, in this sense, that, if their work had been marked by another equally competent set of examiners, the failure of any assigned one of the successful is a very unlikely event, as unlikely as getting a run of six or seven heads or tails at pitch-farthing?"<sup>1</sup> The public examinations in which I have sought the answer to this question fall into three categories: the India Civil Service, the Army, and the Home Civil Service Clerkships of the second order. I find the element of chance in these public examinations to be such that only a fraction—from a third to two-thirds—of the successful candidates can be regarded as quite safe, above the danger of coming out unsuccessful if a set of equally competent examiners happened to be appointed. A corresponding proportion of the successful—from two-thirds to one-third—must be described as unsafe. A rather larger *number*—though much smaller *proportion*—of the unsuccessful candidates would have a chance of succeeding at a re-examination; for instance, at three Clerkships Examinations, 25, 43, 72, out of a total number of unsuccessful, respectively, 126, 102, and 121.

"This question may also be asked: What

<sup>1</sup> Odds of 100 to 1, corresponding to a height above the Honour line of 3.5 times the "probable discrepancy." (Foot-note by Prof. E.)

## EXAMINATIONS—NOTE E

number of displacements is most likely to occur in the event of a re-examination? The answers in some cases which I have scrutinised proves to be about a *seventh part* of the successful. Thus, where fifty clerks were appointed, it may be expected—the most probable event is—that seven would be displaced by a re-examination of the same work."

Prof. E. gives here an elaborate table relating to the results of a number of Civil Service and Army Examinations for which reference must be made to the original.

Prof. E. comes to the conclusion that the distinction between adjacent candidates at such examinations as arranged in accordance with the marks allotted is in hardly any case a real distinction. "Differences of marks on the strength of which one Civil Servant is promoted<sup>1</sup> above his fellow correspond to only the faintest probability of real superiority. Occasionally, but by no means invariably—in examinations decided by an aggregate of marks—the senior or the first two or three break away from their neighbours by an interval which cannot be ascribed to accident. But between the adjacent men among the rank and file the distinction is hardly ever real."

**24. Comments by Prof. E. on his Results.**—Prof. E. thinks his results err, if anything, on the side of moderation, for the following reasons: (1) The correction for mutually compensating errors (see

<sup>1</sup> The author, by "promoted," means here "placed at the entrance examination" above another candidate, being thus entitled to a superior initial position in the service.

## EXAMINATIONS—NOTE E

§ 8) is probably an over-correction, and there is "an intractable minimum of constant error"; (2) the division used for this correction,  $\frac{1}{10}$ , is too great for some examinations, the average number of papers being probably nearer seven than ten; (3) the assumption that differences of *scale* of marking (see § 7) are corrected, or can be allowed for is "precarious," since candidates do not always take the same subjects; (4) owing to changes of *scale*, the hypothesis that the Honour line may be regarded as constant, whatever the change of examiners, does not hold. Hence the error is greater than that shown.

On the other hand, the uncertainty about high marks appears to be less than about lower ones; in order to obtain the best results we should consider only those above the line of what may be regarded as certain failure.<sup>1</sup> "Possibly the uncertainty of some particular examinations may have been exaggerated."

25. *Question of making Marks of a given Examination conform to a given Frequency Curve* [see § 15 (3)].—Prof. E. suggests that any hard-and-fast rule would be dangerous. Referring presumably to Government examinations, he says:

"Perhaps the best general principle is that the 'quartile' and 'median' should be fixed for

<sup>1</sup> It seems clear *a priori* that examiners will not take as much trouble in regard to candidates whom they regard as unlikely to succeed. I do not know whether, at the date of the examinations investigated by Prof. E., the Civil Service Examiners made the "deduction for smattering" of which the existence has long been known, but which was first officially described in the Report of Mr. Stanley Leathes' Committee (see p. 17).

## EXAMINATIONS—NOTE E .

each class of examination by a Committee of permanent officials, in the light of an extensive experience *in pari materia*; that the examiner on each occasion should be expected to furnish a set of marks roughly conforming to the pattern thus prescribed; and that any adjustment which the marks furnished may seem to require, should not be performed without the concurrence of the acting examiner, who has scrutinised the merit of each particular candidate."

c. 26. *Inevitability of an Element of Chance—Suggestions for the Future.*—Professor E. points out that there is an incorrigible minimum of uncertainty, and is satisfied that in some of the cases investigated that minimum could not be substantially reduced by any method of correction.

For hardship entailed on unsuccessful candidates excluded by chance he says: "Perhaps here, as in other grievances, mere publicity is a sufficient, and the sole obtainable, remedy. The general recognition of the element of chance in examinations would mitigate the disappointment of those who, for want of a few marks, are either altogether excluded from the class of the successful, or compelled without preferment to take a low place in that class."

Prof. E. suggests, rather for the sake of facilitating the conception of his theory "than as a serious practical proposal," that candidates within the limits of uncertainty should draw lots for the vacant places allottable, each candidate receiving a number of tickets proportional to the chance that he is really deserving, as accurately ascertained by the theory of probabilities. He points out that such a proposal would

## EXAMINATIONS—NOTE E

mitigate the sense of injustice in disappointed individuals and "do more to bring home to the public mind the aleatory character of examinations than even an article in the *Journal of the Royal Statistical Society*."

Prof. E. points out the element of certainty as well as that of chance in public examinations. He compares their results to those of a barometer uncorrected for violent changes of temperature. "That examinations are a very rough test of merit, but they do serve to distinguish excellence in certain qualities, is a conclusion which recommends itself to practical and sensible persons."

But "mere common sense, without the calculus of probabilities, is not competent adequately to define the limits of chance and certainty. Common sense might be embarrassed in defending its position against the awkward fact which has been adduced, that even in the best regulated examinations one examiner occasionally differs from another to the extent of 50 per cent.; for common sense cannot fully appreciate that mutual compensation of errors in virtue of which a peculiar stability attaches to aggregates of marks."

In a charming peroration (so suave one hardly dare suspect it of irony) Prof. E. suggests that "To examiners at least it will be interesting to test the accuracy of the instrument with which they work. The statistical study may beguile the monotony of their task."

**27. Conclusion.**—In presenting the above summary of Professor Edgeworth's papers I must disclaim once more any pretension to be an authority on statistics



## EXAMINATIONS—NOTE E

or to add by quotation any weight to Prof. Edgeworth's results. They are those of a recognised expert to whom one cannot but be grateful. In spite of the soothing remarks which I have, I hope, reproduced as faithfully as the disturbing figures, the papers leave me, and I believe they will leave others, with the conviction that the necessity for further investigation is real and pressing. Prof. Edgeworth's results, significant as they are, are based on data some thirty years old. What are the facts now? Ought not each examination on a large scale to be periodically subjected to the criticism of a professional statistician? We cannot avoid the effects of chance in examinations—such chance exists in all human affairs—but it is surely a duty both to ascertain what they are, and to reduce them to a minimum.

And I repeat once more my belief that, if examiners asked themselves exactly how they could define in words what the passing of an examination means, so as to be able to give a definite assurance to the public in regard to what all the successful candidates can do, the element of chance would be singularly reduced.

I make one last quotation from Prof. Edgeworth, which is appropriate here though he uses it in a somewhat different connection :

“What should we think of a company who, dealing in thousands and millions sterling, grudged the payment of a professional accountant to check the accounts of their Directors? It would be a similarly false economy on the part of educated society, considering the stupendous sum of mental labour involved in the system

## EXAMINATIONS—NOTE E

of competitive examinations, to grudge the expenditure of a little additional effort on a technique which professes to bring to book, and test the figures of the examining bodies who are responsible for the direction of that labour."

## NOTE F

### ON KNOWLEDGE-TESTS AND MEMORY-TESTS

It is frequently impossible in a written examination to distinguish knowledge, implying understanding, from pure memory. The instance of the candidate who, in translating a passage from a prepared book, carries his translation beyond the passage set, illustrates the point.

It has been suggested by a critic that I have overestimated the importance attached to memory-tests in examinations at the present time. I may point out in this connection that Prof. Burnet, in his *Higher Education and the War*, strongly advocates the use of set books, and deprecates the use of "unseen translation" in classical teaching in schools. I am not prepared to contest the soundness of his views in this particular matter; I quote them as showing that memory-tests are not obsolete, and that precautions should be taken in using them.

Since the publication of the above criticism, one extremely successful candidate told me that, having to pass examinations in English (which, although he is a British subject, is not his mother-tongue) he learnt off by heart answers which he did not understand to questions which he recognised by catch-words, and passed a series of examinations in this way. He assured me that his case was an extremely common one in his own part of the British dominions. An

## EXAMINATIONS—NOTE F

Englishwoman who was first or second in each of the four subjects of a first-year University course in natural science at one of the largest University colleges in Great Britain a few years since, has also assured me that in physics, in which she took the second place, she would have been quite incapable of explaining some of her own answers; and that her personal success gave her a rooted disbelief in the results of examinations as now conducted.

Charles Lister, while serving his diplomatic apprenticeship at Rome in 1912, commented on "the blissful ignorance of commerce" in the diplomatic service (now, about to be remedied), and wrote: "I should suggest some compulsory subject in the examinations in banking, finance, and trade, but probably some Hawkins would come along and put it all into a book in the form most easily learnt and most easily forgotten, and little would come of it" (*Charles Lister: Letters and Recollections*, 1917, p. 45). In the considerable literature on examinations the views of the recently examined are comparatively rare; I attach importance to them, especially in the case of able candidates such as those above cited. My critic, M. H. (see *Times Educational Supplement* for August 2, 16, and 23, 1917) writes: "Only after-life can show how far the acquirements of memory are fleeting: unconscious stores may play their part in the intellect; it is better to have learned and lost than never to have learned." The excellent epigram seems to me applicable only to knowledge which has been assimilated. The examples quoted above show that unassimilated knowledge may and does still serve its turn in the examination-room.

## NOTE G

### ON THE USE OF THE TERM "ORDER OF MERIT"

I HAVE consulted my friend Prof. S. Alexander, on the significance attached to the word "merit," as used by professional writers on ethics; he has kindly referred me to Leslie Stephen's *Science of Ethics*, ch. vii., § 1 Merit, and § 2 Effort; James Martineau's *Types of Ethical Theory*, Vol. II., Book I, ch. iii., Merit and Demerit; and to his own *Moral Order and Progress*. In the last-named work (p. 194), Prof. Alexander writes of merit: "Though in reality a simple idea, it has given rise to the greatest controversy, from the apparent divergence with which it is applied." This is not the place to discuss general theories of ethics, even, if the writer were competent to do so. But the objection to the use of the word "merit" in the phrase "order of merit" is that it really leads to confusion of ideas, as indicated in § 17, p. 63 above. I find myself in modest disagreement with Prof. Alexander's reconciliation of different views, and in general sympathy with Martineau's interpretation, which Prof. Alexander thinks is also that of "the man in the street."

I think that confusion arises frequently because in using the word one is apt, when speaking of an

## EXAMINATIONS—NOTE G

action, to lose the distinction between (a) the intrinsic merit of the action considered independently of the author, and (b) the merit involved in its performance. The more dazzling the action, the more difficult does the public find it to make the distinction. A chance observation of scientific importance may confer undying fame on the observer. Heroism may confer immortality on a failure.

But the reality of the distinction will be made clear by a concrete hypothetical example connected with our subject. Suppose you are shown a short piece of French prose, irreproachable in grammar and "quite good" in style, of which you do not know the author. You say, "This piece has merit." You then ask by whom it was written, and are told that it was written (say) by Anatole France. "Ah," you say, "not up to his usual level." You are then informed that the previous information was mistaken, and that the piece was by an English boy who has never been in France; and you are filled with admiration. Clearly, one attributes great merit to the English schoolboy for a performance for which one would attribute none to Anatole France, yet the value of the piece of prose itself remains the same.

The example is not so far removed from reality as it may seem. While it is quite true that French authors do not enter for our examinations in French, candidates who have been partly trained in France do so, and compete against others who have contended against the greatest possible difficulties, it may be both of natural disposition and of external opportunity, in order to achieve the same result as those more favoured by nature and by fortune

## EXAMINATIONS—NOTE G

Clearly we have no right to say, then, that examinations place the candidates *themselves* in "order of merit." But we ought to be justified in saying that examiners place candidates in the order of proficiency shown by their answers, which is another matter.

To the philosophical critic, I venture to point out that the case is simplified by the fact that examination answers are not (in general) productions so dazzling that all distinction between the merit of the author and of the production is perforce obliterated in the public eye and understanding. Their intrinsic public value, except in a few delightful instances, is nought. Having served their purpose, they pass, unregretted, to the oblivion of the paper-mill. •

## INDEX

- Ability, tests of 27, 29, 84, 88  
*Action of Examinations*, Iatham's, 13  
 Actuaries, examinations for, 65, 70  
 Adams, Prof. J., 48  
 Adelong, J. C., 95, 96  
 Administration of examinations, 22, 78-9  
 Administrators, association of, with examiners, 42  
 Aggregates of marks, 16, 76, 117, 123-5  
 Alexander, Prof. S., *Moral Order and Progress*, on word "merit," 138  
 Anonymity of candidates, 63, 74  
 Answers, non-representative, error due to, 113  
 Apperception, 93  
 Army examinations, 130  
 Arnold, Matthew, definition of culture, 8, 94  
 "Arts" students, to attend classes in natural science, xvi  
 Arts, Royal Society of:  
   lecture at, ix, 1-34  
   speech by Lord Cromer at, 34-48  
 Assistant examiners, 114-17
- Beaulieu, Leroy, 35-6  
 Benson, E. F., 94  
 Bernhardt, 100-101  
 Bonasse, Henri, 94  
 Bowley, A. C., *Elements of Statistics*, 104  
 Browne, Prof. E. G., 28  
 Bryan Prof., 19, 61
- Bryant, Mrs., results of examining same papers at different times, 124  
 Burnet, Prof., *Higher Education and the War*, xiv, xv, 6, 136
- Cambridge Classical Tripos:  
   details of, 125-6  
   statistics of, 112  
 Capacity-tests, 11-15, 17, 20-3, 52, 62-3, 67, 84, 91  
 Chance in examinations, 29, 31, 59, 65-6, 102-35  
 Character, examinations as tests of, xi, 28, 73, 76, 88  
 Characteristics:  
   mental, variability of, 55  
   tests of, 53-6, 64-5, 67  
 Chatham, Lord, 42  
 Chinese Examinations, 4  
 Chirol, Sir Valentine, 28, 29, 41  
 Civil Service Commission (Fourth Report), 1914, ix, xii, 80-8  
 Civil Service Examinations, ix-xii, 7, 27, 28, 59, 67, 75-7, 80-8, 89-92, 109, 129-32  
   Class I, 67, 75-7, 89, 109  
   Report of Treasury Committee on, xii, 89, 131  
 Classes, indicated by letters, 19, 20, 61, 62, 108  
 Classical teaching, examinations as test of, xiv  
 Classification of examinations, 11-12, 51-3  
 College of Preceptors, 7  
 Common sense, examinations as test of, 77, 84-5



# INDEX

- Compensation of independent errors in marking, 104, 130-1.
- Competency of examiners, 34, 103.
- Competition of subjects, 109.
- Competitive examinations, ix-xii, 6, 7, 20, 27, 28, 32, 37, 39-42, 46, 47, 49, 62, 64, 66, 80-92 (*see also* Civil Service Examinations).
- Composition, need for training in, 72.
- Concentration, effect of examinations on powers of, 71.
- Condition of candidates during examination, 64, 67, 106.
- Conduct of examinations, 22, 78-9.
- Consultative Committee of Board of Education, 30.
- Cramming, xii, 5, 82-3.
- Criteria of tests, 53-6.
- Cromer, Lord:  
method of selection, 28-9, 45-6, 75.  
on inquiry into examinations, ix.  
speech by, 34-48.
- Culture:  
and *Kultur*, xiv, 95-101.  
effect on education of examinations designed to test, 73.  
examinations as test of, xv-xviii, 8, 10, 49.  
meaning of, xiv-xviii, 8, 9, 93-101.
- Culture subjects, xv-xviii, 25-6.
- Curriculum:  
effect of examinations on (*see* Education).  
School, interruption of, by examinations, 83.
- Definition of examinations, 4, 49.
- Differences of marking between different examiners, 16-17, 118-23, 128.
- Difficulty of examinations, 7-8.
- Dispersion of marks, 118.
- Edgeworth, Prof. F. Y., xviii, 17, 29, 59, 102-35.
- Education:  
effect of examinations on, xv-xviii, 2-6, 9, 13, 23-5, 32-4, 69-74, 82-3.  
Examinations as test of, 91 (*see also* Culture).
- Education Reform Council, xviii.
- Effect of examinations:  
on candidates, 22-4.  
on characteristics, 54.  
on concentration, 72.  
on education (*see* Education).  
on health, 23, 24, 29, 31-2, 67-8, 74.  
on powers of exposition, 71-2.  
on Secondary Schools and Universities, xiii.
- Egyptian Services, method of appointment for, 28-9, 45-6, 75.
- English composition, details of examination in, 116-7.
- English language, examination in, in Treasury Committee's Scheme, 90-92.
- Entrance Examinations, University, 50.
- Errors in marking, 57-63, 102-35.  
Summaries of, 111, 128.
- Examen*, 53.
- Experimental Psychological Tests, 53, 77-8.
- Exposition, effect of examinations on powers of, 71-2.  
*viva voce*, as test of, 68.
- External examiners, 52, 66, 69.
- Fairness of examinations, 74-5.
- Fatigue:  
of candidates in examinations, 23, 24, 29, 32, 67-8.  
of examiners, errors due to, 111.
- Fisher, H. L., xiii, xviii.
- Frequency Curve, 105, 107, 117-18, 131-2.
- Galton, Sir Francis, testing general ability, 29.
- General ability, 27, 29.
- General impression, marks for, 61.
- Germany, examinations in, 5-6.

# INDEX

Gilda, 4  
 Goethe, 93-4, 97-9  
 Greek, examinations in, 9  
 Greek lettersystem of marking,  
 19, 20, 62, 108  
 H., M., 137  
 Hardinge, Lord, 37  
 Harnack, 99, 100  
 Hartog, P. J.:  
     *Aims of the New Curricula in*  
     *French Schools for Boys*, 92  
 British Association paper on  
     *Examinations*, 77  
*Encyclopædia Britannica* article  
     on "Examinations,"  
     xviii  
     *Writing of English*, 72, 77, 94  
 Hayward, Dr. F. H., 77  
 Health (see Effect of Examinations)  
 Herbert, Auberon, *Sacrifice of*  
     *Education to Examinations*, 69  
 Hinneberg, *Die Kultur der*  
     *Gegenwart*, 99  
 History:  
     examinations in, 12, 25, 66,  
     73  
     teaching of, 25-6  
 Honours:  
     line, 124-5, 129  
     lists, 114-15  
     men, competition of, *inter se*,  
     84-5  
 Huxley, *Science and Education*,  
     viii  
 Hygiene of examination-room,  
     31-2  
 Ideal typical examination:  
     choice of, for investigation,  
     105-6  
     error of marks in, 111-14  
 Idiosyncrasies of candidates, 21  
 examiners, 21, 111, 118 (see  
     also Personal Equation)  
 Impartiality of examinations,  
     74-5  
 Impression, general, marks for,  
     61  
 Instructions to candidates:  
     need for definite, 78-9  
     in scheme of Treasury Com-  
     mittee, 92

Investigation on examinations,  
     need for, i-xii, xvii, 18,  
     30-34, 41, 60, 79, 87  
 Judgment, use of examinations  
     to test, 77, 84-5  
 Keatinge, M. W., *Studies in*  
     *History*, 13  
 Knowledge-tests, 11-14, 21-3,  
     52, 84, 85, 136-7  
 Kultur and Culture, 95-101  
*Kultur and Humanism*, Prof.  
     Barnet, xiv  
 "Laboratory Science," 71  
 Latham, H., *Action of Examinations*, 13  
 Latin composition, marking of,  
     by different examiners,  
     119-123  
 Leathes, Stanley, *What is*  
     *Education?* 61, 69 (see  
     also Civil Service Ex-  
     amination — Class I,  
     Report of Treasury Com-  
     mittee)  
*Licentia docendi*, 5  
 Lippmann, Prof. Gabriel, 15  
 Lister, Charles, 137  
 Literature:  
     as culture subject, 26  
     examinations as test of know-  
     ledge of, 12, 66, 73  
 Livingstone, R. W., *Defence of*  
     *Classical Education*, xvi  
 Macdonnell, Lord, Royal Com-  
     mission presided over  
     by, ix-xii, 80-8  
 Magnitude of examination  
     operations, 33, 50  
 Mark, true, 102  
 Marking:  
     examiners not to have ex-  
     cessive amount of, 79  
     of Latin composition by  
     different examiners, 119-  
     123  
     problems of, 31-2, 57-64  
 Marks:  
     aggregates of, 16, 76, 117  
     123-5

# INDEX

## Marks (cont.):

- differences of, between different examiners, 102-11, 118-23, 128
- different results in, obtained by marking the same work at different times, 124
- dispersion of, 118
- errors in:
  - compensation of, 104
  - due to haste, 110-11, 116-17
  - in ideal typical examination, 119
  - for general impression, 61
  - negative, 61
  - statistics of, 103
  - Treasury Committee's scheme of, 90-91
- Martineau, James, *Types of Ethical Theory*, on word "merit," 138
- Mathematics, examinations in, 19
- Matriculation examinations, 50
- Maxima, absolute, 61
- Mediæval examinations, 4-5, 49
- Medical tests, 53, 78
- Melbourne, Lord, 38
- Memory-tests, 11-14, 17, 20, 22, 52, 62-3, 67, 74, 136-7
- Merit, Order of, 10, 16, 49, 58, 62-3, 106, 138-40
- Métrel*, by Prof. Edgeworth, 102
- Meyer, *Konversationslexicon*, 96
- Michaelis, Paul, 100
- Miers, Sir Henry, 48
- Minimum sensible*, 106, 108, 110, 111
- Mistake, single, varying importance attached to by examiners, 122
- Muench, W., 99
- Münsterberg, 77
- Natural ability, 84, 88
- Negative marks, 61
- Negligence of haste, errors due to, 110, 111, 116, 117
- Nervousness in examinations, 20, 21, 68
- Newman, John Henry, 8, 93
- Nomination for appointments, 37-8
- Nunn, Prof. J. P., xviii, 71

Objects of examination, 3-11, 17, 32, 49-51, 56-7, 69-71, 79, 103

Order:
 

- in Council of 1870, 47
- of Merit, 10, 16, 49, 58, 62-3, 106, 138-40

Origin of examinations, 4, 49-51

Oxford *Literæ Humaniores*, examination in Philosophy details in regard to, 126

Passes, causes of variation of, 60

Passing an examination, meaning of, xiii, xiv, 17, 22, 27, 55-7, 59, 60, 79, 134-5

Pattison, Mark, 13

Paulsen, F., 5, 6, 99

Personal equation, errors due to, 107-11, 117

Poor Law Commission, on popular election, 36

Practical examinations, 11, 52, 58-9

marking in, 20

Probabilities, problems of, 102-4

Professional examinations, 2-3, 9, 50, 56, 59, 65-6, 79

Progress, examinations as tests of, 62-3

Psychological tests, 53, 77-8

Psychology of examinations, 32

Purposes of examinations, 3-11, 17, 32, 49-51, 56-7, 69-71, 79, 103

Qualities which examinations can test, 10

Rashdall, Hastings, *Universities of Europe in the Middle Ages*, 6

Reay, Lord, Committee on Oriental Studies, 28

Reports on candidates, 64

Rickert, H., *Kulturwissenschaft und Naturwissenschaft*, 96

Rosebery, Lord, 34, 42

Royal Commission:
 

- on Civil Service (Fourth Report), 1914, ix-xii, 80-8
- on Examinations, plea for, 30-4, 47-8

# INDEX

- Sadler, M. E., xiii, 30, 48  
 Sample, problem of examination, 64-9  
 Schiller, 97, 99  
 School-leaving certificate as condition for clerical examinations, 84  
 School-leaving examinations, 50  
 Schools:  
   Examinations as test of, 7, 8, 10, 49, 69  
   Secondary, Education Council, xvii  
     Secondary, effect of examinations on, xiii  
 Schuster, Prof. Arthur, 48, 115  
 Science:  
   aesthetic and individual side in, xvii  
   greater importance attached to, in Treasury Committee's scheme, 91-2  
   speculative, details of examination in, 126-7  
   students to attend classes in literature and history, xvi  
   view of, as unimaginative, xvi  
 Selborne, Lord, method of choosing candidates, 29  
 Selection:  
   appointment by, 28-9, 42-6, 75  
   by board, 46, 75  
 Set books, examination on, 136  
 Shaw, Sir Napier, 13, 72  
 "Smattering, deduction for," 131  
 Smattering of knowledge, 19  
 Soudanese Civil Service, 28-9, 45-6, 75  
 Spearman, Prof. C., 77  
 Standard, 17-19, 31-2, 58, 102  
 Statistics of:  
   Cambridge Classical Tripos, 112  
   examinations, Prof. Edgeworth's investigations on, 102-35  
   marks, 103  
 Stephen, Leslie, *Science of Ethics* on word "merit," 138  
 Stimulus, examinations as, 71  
 Strain of examinations, 23-4, 29, 31-2, 69, 74  
 Subject, examining in a, 1  
 Subjects, competition of, 109  
 Taylor, F. W., 77  
 Teachers as examiners, 32, 70-1  
 Terms of reference, suggested for Royal Commission, 31  
 Time, effect of examinations by over-absorption of, 73-4  
*Times Educational Supplement*, xii  
 Training of Examiners, 34  
 "Traps," 79  
 Treasury Committee on Oriental Studies, 28  
 Trustworthiness of tests, 34  
 Tylor, *Primitive Culture*, 96-7  
 Typical Examination:  
   choice of, for investigation, 105  
   total error of marks in, 113  
 "Unit Knowledge," no such thing as, 54-5  
 Universities, effect of examinations on, xiii  
 University Examinations, xiii, 6-8, 52  
   Cambridge, 112, 125-6  
   entrance, 50  
   Oxford, 126  
 Unseen translations, 136  
 Value of examinations, general question of, 69-71 (*see also* Effect of Examinations)  
 Variations of passes, causes of, 60  
 Venn, Dr., 109  
 Viva-voce examinations, 11, 20-22, 52, 68, 70, 74-5, 85-6, 91-2  
   by committee of three, 86  
   in Treasury Committee's scheme, 81  
   public, 92  
   "Vocational bureaux," 78  
 Watson, Arthur, xviii  
 Whitehead, Prof. A. N., xiv-xv, xviii  
 Women, effect of examinations on, 24, 74

PRINTED IN GREAT BRITAIN  
BY HAZELL, WATSON AND VINEY, LD.,  
LONDON AND AYLESBURY.





